

[11] Japanese Registered Utility Model No. 3060617
[24] Issue Date: June 16, 1999
[21] Japanese Utility Model Application No. 10-10405
[22] Filing Date: December 28, 1998
[72] Devisor: Akira INOUE
[73] Proprietor: Hitachi Software Engineering Co., Ltd.
[54] Title of the Device: IMAGE COMPOSITION OUTPUT UNIT
WITH A SPEECH SYNTHESIS OUTPUT FUNCTION

*** NOTICES ***

JPO and NCIPJ are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.**** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

CLAIMS

[Utility model registration claim]

[Claim 1] A photography means to photo a photographic subject, and a sound recording means to record the voice which the photographic subject generated, A photography image data storage means to memorize the photography image data photoed with said photography means, A sound recording voice data storage means to memorize the sound recording voice data recorded with said sound recording means, An image pattern data storage means by which said photography image data and two or more image pattern data to compound were memorized, A sound pattern data storage means by which said sound recording voice data and two or more sound pattern data to compound were memorized, An image pattern data selection means to choose desired image pattern data among two or more image pattern data, An image composition output means to compound the image pattern data chosen as said photography image data with said image pattern data selection means, and to output the image data after composition, A voice pattern data selection means to choose desired sound pattern data among two or more sound pattern data, It has a speech synthesis output means to compound the sound pattern data chosen as said sound recording voice data with said sound pattern data selection means, and to output the voice data after composition. Furthermore, a storing means to store in a pocket storage the display playback program for displaying the printing means which carries out the printout of the image data after said composition to a seal, the image data after said composition, the voice data after said composition, and the image data after said composition, and reproducing the voice data after said composition, A distribution means to distribute the display playback program for displaying the image data after said composition, the voice data after said composition, and the image data after said composition, and reproducing the voice data after said composition to the mail address specified by a user by the network course, The image composition output unit with a speech synthesis output function equipped with any one ** or plurality.

[Claim 2] The image composition output unit with a speech synthesis output function according to claim 1 characterized by changing said sound recording voice data into alphabetic data, compounding the changed alphabetic data to image pattern data, and having further a voice data conversion means to store in said image pattern data storage means.

[Claim 3] The image composition output unit with a speech synthesis output function according to claim 1 which reads the image data, the sound data, or both which were beforehand memorized in said pocket storage, and is characterized by having further a pocket storage reading input means to use sound data as sound recording voice data or sound pattern data, and to input them, respectively by using image data as said photography image data or image pattern data.

[Claim 4] It is the image composition output unit with a speech synthesis output function according to claim 1 characterized by equipping sound data with a data reception input means to input as sound recording voice data or sound pattern data, respectively, further by receiving the image data, the sound data, or both via a network, and using the image data which received as said photography image data or image pattern data.

[Translation done.]

* NOTICES *

JPO and NCIPi are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.**** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed explanation of a design]

[0001]

[The technical field to which a design belongs]

This design is installed in an amusement park, a game center, etc., compounds and photos a photographic subject (user) and the image pattern prepared beforehand, starts the equipment which prints out the photography image as a seal, a business card, etc., and is related with the image composition output unit with a speech-synthesis output function whose output compounded not only a photography image but sounds, such as voice, and was enabled from the predetermined output media in detail.

[0002]

[Description of the Prior Art]

A user's portrait image is photoed in recent years, the image data specified as the photoed portrait image because a user specifies image data, such as a background prepared beforehand, is compounded, and the image printing equipment which prints a synthetic result to a print sheet / seal, a business card, etc. is installed in the amusement park, the game center, etc. This image printing equipment is equipped with the monitor which displays the memory and the photography image which memorized beforehand two or more image data for composition, such as a camera for photoing the user who is a photographic subject, and a background, the image composition means and photography image which compound image data for composition, such as a background, and the image data for composition, the printer which prints a synthetic image on a seal, a business card, etc. as a fundamental configuration.

[0003]

[Problem(s) to be Solved by the Device]

By the way, it is clear from the above-mentioned configuration for attaining the above-mentioned object of printing the created synthetic image on a seal, a business card, etc., and its object that the above-mentioned conventional image printing equipment deals with only image data. The camera, the monitor, an image composition means, a printer, etc. are constituted from the conventional image printing equipment by only the means for dealing with an image, and sound-collecting which incorporates voice, music, etc. in equipment, and a means to perform processing about a sound recorder style or these voice were not established. That is, since the conventional image printing equipment had not made the information about sounds, such as voice, music, and a sound effect, applicable to utilization, it had the problem that voice data, such as background sounds, such as a user's voice recorded by arbitration and a natural sound, and background music, could not be dealt with together with an image.

[0004]

Moreover, in the conventional image printing equipment, an output is only two-dimensional expansion of printing in a print sheet / seal, a picture postcard, etc., and it was not able to be said that the synthetic image was edited and processed further, or a synthetic image was re-printed like an extra copy after printing termination of a synthetic image. Moreover, the conventional image printing equipment was only used for rewriting of a background pattern, or maintenance of equipment, even if it connected with the network.

[0005]

By adding the function in which voice data, such as background sounds, such as a user's recorded voice and a natural sound, and background music, can be treated to the conventional image printing equipment, as the object of this design can deal with both an image and voice, it is from both an eye and a lug to provide a user with the image composition output unit with a speech synthesis output function which can give presence.

[0006]

Moreover, other objects of this design deal with an image and voice as digital data. In addition to the printing means to a print sheet / seal, as an output means by having a storing means to various information storage media, and a distribution means to the assignment mail address which went via the network It is in offering an image composition output unit with a speech synthesis output function with which it also becomes possible to enable it to perform display and playback of image data and voice data behind, and for a user to edit and process voice freely in the image.

[0007]

[Means for Solving the Problem]

The object of this design is attained by preparing the sound composition output which compounds and outputs the sound data which a user chose among two or more sound data memorized by the sound-recording means records a user's voice and memory is made a means to memorize, a sound data-storage means to by which two or more sound data, such as a background sound and background music, were memorized, and this sound data-storage means, and the voice data of the user who recorded with said sound-recording means.

Other objects of this design Moreover, the output means to various storages, such as an IC card and a floppy disk, The distribution means to the assignment mail address via a network, the input means from various storages, The display means to the monitor of image data and the playback means from the loudspeaker of voice data are used. The compound image data and sound data are distributed to various storages via an output or a network to an assignment mail address, the image data and sound data which were outputted to various storages later are inputted, and it is attained by reproducing from a display and a loudspeaker to a monitor.

[0008]

[The gestalt of implementation of a design]

Hereafter, the gestalt of operation of this design is explained to a detail with reference to a drawing.

Drawing 1 is hardware configuration drawing showing 1 operation gestalt of the image composition output unit with a speech synthesis output function of this design. even if a monitor-cum-these touch panels 1, the microphone 2, and the loudspeaker 3 be being fix to the revolution base 4 which pick up a user (photographic subject) by the image composition output unit with a speech synthesis output function of this operation gestalt and the revolution base 4 rotate in it, a user look at a monitor in it by be form a monitor-cum-the touch panel 1, a microphone 2, and a loudspeaker 3, a touch panel 1 operate in it, and its voice record in it, and the playback sound of guidance or voice data can ask it now. Moreover, the revolution base 4 rotates with a driving gear 5, and a camera 6 can be constituted so that it may move up and down with a driving gear 7, and it can photo a user now in free composition. In this case, the visual field of a camera 6 is set up possible [photography of the whole body image from a user's head to a step]. The processor 8 which performs an image processing etc. is formed in the interior of the equipment which does not appear from a user. This processor 8 can be distributed now to the mail address to which it connected with the network 9 and the user specified the created image data and voice data as it via the network 9. In the seal printing machine 10, the printout of the image data and voice data which set to the processor 8, and were processed and compounded is carried out to a seal. Moreover, in the READ/WRITE equipment 11 of media, such as an IC card, image data and voice data can also be outputted to a pocket storage like an IC card.

[0009]

Drawing 2 is the block diagram showing the important section of the configuration of 1 operation

gestalt of this design.

CPU201 of the image composition output unit with a speech synthesis output function concerning this design is connected to memory 202, the I/O-hardware-control section 203, an image and the voice input section 204, the pattern memory 205, an image and the speech processing section 206, an image and the voice output section 207, the data storage section 208, and the data input section 209. Each circuit block is formed on a semi-conductor substrate or a glass epoxy group plate of a well-known semiconductor integrated circuit applied technology, for example.

CPU201 receives the compression digital signal from digital camera I/F211 and microphone I/F212, and compounds it by the program in which the compression digital signal is stored by an image and the speech processing section 206, and the printed output processing by the video printer 217 (equivalent to the seal printing machine 10 of drawing 1) and the output to the information storage by READ/WRITE equipment (equivalent to the READ/WRITE equipment 11 of the IC card of drawing 1) 218 are performed.

[0010]

Memory 202 memorizes the image data and voice data which were incorporated from an image and the voice input section 204, and also has memorized as a program the instruction group used as the instruction emitted to CPU201, and the control procedure of an image composition output unit with this speech synthesis output function. Moreover, the processed data of CPU201 are also temporarily memorized as work-piece memory.

An image and the voice input section 204 change into a digital signal the image data which the camera (equivalent to the camera 6 of drawing 1) 220 photoed, and the voice data which the microphone (equivalent to the microphone 2 of drawing 1) 221 recorded from an analog signal with A/D-conversion equipment (not shown).

At this time, it is also possible about image photography to perform chroma-key processing etc. The pattern memory 205 memorizes image pattern data, such as a background for piling up with the image and voice data in memory 202, and a character, and sound pattern data, such as music and voice of a character. It is also possible to store external data so that it may mention later, and to store the complex data itself stored in memory 202 as well as the ability to use as image pattern data or voice pattern data as image pattern data or sound pattern data.

[0011]

An image and the speech processing section 206 perform synthetic processing of an image and voice based on the image and voice data of memory 202, and image pattern data and sound pattern data of the pattern memory 205.

An image and the voice output section 207 output the data which an image and the speech processing section 206 processed. It has connected with a video printer 217 and READ/WRITE equipment 218, and an image and the voice output section 207 are outputted to a video printer 217 or READ/WRITE equipment 218 according to directions of a user or the gestalt of an output.

The data storage section 208 rewrites the content of the pattern memory 205. It collates with the date of the data currently written in the pattern memory 205, and if the data memorized by the data storage section 208 are new, the data in the pattern memory 205 will be rewritten. Thereby, the data memorized by the data storage section 208 can be used as image pattern data and sound pattern data, such as a background, voice, and a character.

[0012]

It connects with READ/WRITE equipment 218 and the data input section 209 incorporates image data, sound data, etc. of the pocket storage 219 with digital information. The data input section 209 once stores the incorporated data in the data storage section 208. The data storage section 208 will rewrite the data in the pattern memory 205 to the data of the data storage section 208, if the data is judged to be a thing newer than the data of the pattern memory 205 as compared with the date of the data in which it was written by the pattern memory 205. Thereby, the data which the data input section 209 incorporated can be used as image pattern data and sound pattern data, such as a background, voice, and a character. Furthermore, it is also possible for the data input section 209 to store the incorporated data in memory 202, to overwrite the data,

and to compound.

[0013]

By the way, it connects with communications control I/F210, digital camera I/F211, microphone I/F212, a sensor 213, a touch panel 214, the coin control section 215, and a timer 216, and the I/O-hardware-control section 203 controls these whole.

Communications control I/F210 is used for connection with an external network.

Communications control I/F210 is connected to the device management server which is outside through a dedicated line. This device management server (not shown) aims at managing the existence of failure. Moreover, a new background pattern, a sound pattern, a character, etc. are transmitted from a device management server, and it memorizes in the data storage section 208 through the I/O-hardware-control section 203. Moreover, communications control I/F210 is connected to optical-communication control I/F223. Optical-communication control I/F223 controls control of Ir-DA which performs infrared ray communication, and the communication link using an optical fiber. The exchange of binary data, such as a program, can be performed by using this interface, direct continuation can be carried out to information personal digital assistants, such as PDA, or a notebook computer, and an information personal digital assistant and a personal computer can incorporate the program and data of an image composition output unit with a speech synthesis output function of this design. Moreover, data and the program which were recorded and created with the personal computer can be sent to an image composition output unit with this speech synthesis output function, and a background pattern, a sound pattern, and a character can be changed into arbitration.

[0014]

The user who exists [whether a sensor 213 detects the existence or nonexistence of the user like a control unit by the infrared sensor, and its user is in a position after the coin charge and] again checks whether it is an unit or it is plurality.

A touch panel (equivalent to the touch panel section of a monitor-cum-the touch panel 1 of drawing 1) 214 displays for example, an actuation menu, and receives the input of various selection conditions. Moreover, a touch panel 214 is CPU's201 detecting resistance when the actuation menu displayed, for example is inputted by the user through A/D-conversion equipment, and computing the coordinate on a screen, and judges which carbon button was chosen by the user.

The coin control section 215 detects the existence of the charge of coin, the charge of inaccurate coin, etc. by CPU201. The configuration and weight of the thrown-in coin are measured, the pattern of the front flesh side of coin is further judged to the coin which is not the Shinsei coin, and it distinguishes whether the thrown-in coin is the Shinsei coin, inaccurate coin, and forged coin. Although the coin control section 215 receives a photography tariff, it is also possible for a tariff acceptance means not to be restricted to a coin detector, but to use a bill reading machine, a prepaid card reading machine, a credit card, etc.

A timer 216 is formed in the body of equipment, when there is nothing by the environment connectable with the management server which is outside by the dedicated line, and it operates with the instruction of CPU201.

[0015]

By the way, after the coin charge, when a touch panel 214 receives the photography directions to a user, a camera (equivalent to the camera 6 of drawing 1) 220 performs a user's photography. As a picture input device, MOS, the usual video camera, etc. may be used for a camera 220 other than a CCD camera. AE (auto iris) function which a camera 220 contains can carry out exposure adjustment, and the image data which the camera 220 photoed can also be memorized in memory 202.

Moreover, the digital still camera which has for example, an Ir-DA optical-communication function as a picture input device may be used. Then, a user can be made to choose the icon used as the selection condition of a menu image by the buffer memory which can store the Ir-DA optical-communication function and the image data of two or more sheets which can receive and transmit the image data photoed beforehand, and the still picture of two or more sheets which carried out JPEG compression can be easily compounded now.

[0016]

On the other hand, the command which records voice by the user, the command which records a background sound are chosen, and sound recording of voice, a background sound, etc. is performed when a touch panel receives the command. As an audio input unit, a recorder, a loudspeaker, etc. linked to LINEOUT222 for external instrument connection other than a microphone 221 are sufficient.

Speaking of an output, when an image and the voice output section 207 output data to a video printer 217, a video printer 217 prints and outputs a color picture to a postcard or a seal. When it outputs to READ/WRITE equipment 218, READ/WRITE equipment 218 memorizes the image data and sound data which were generated to the pocket storage 219 with digital information.

[0017]

It seems that in addition, image data and voice data are distinguished and dealt with in practice, and sound data are not compounded to image data in the above-mentioned explanation like an image and the voice input section 204, or an image and the speech processing section 205 although it is dealt with without distinction of image data and sound data for convenience. It is only sound data which are compoundable to image data and sound data, and after changing into image data to compound sound data to image data, it performs that it is compoundable to image data. The data conversion feature in that case may be prepared as a way stage which should just prepare for the image and the speech processing means 205, and became independent.

[0018]

Moreover, although the control section of the driving gears 5 and 7 of drawing 1 is not shown in drawing 2, actuation control of the driving gears 5 and 7 of drawing 1 is performed through each actuation control section connected to the I/O-hardware-control section 203. Moreover, revolution initiation and a halt are directed on the revolution base 4, when a user operates the foot switch formed for example, in the revolution base floor line. It is directed when a user operates the directions switch similarly formed in the touch panel 1, or the foot switch formed in the revolution base floor line also about vertical movement of a camera 6.

[0019]

Drawing 3 is the outline perspective view showing a concrete configuration in case the image composition output unit with a speech synthesis output function concerning this design outputs to a pocket storage. The gate 301 for taking storages, such as an IC card, in and out is established in the actuation side side of a case 300 as shown by a diagram. Thereby, the READ/WRITE equipment 11 shown by drawing 1 can perform I/O of the image data in a pocket storage, or sound data.

[0020]

Drawing 4 is drawing showing the structure and the output destination change of output data of an image composition output unit with a speech synthesis output function. Output data 400 are roughly divided into two, the program section 410 and data division 420. The program section 410 has the image data display function 411 and the voice data regenerative function 412, and data division 420 are divided into two, the image data 421 after composition, and the voice data 422 after composition. Output data 400 have the program section 410, because display playback is enabled for data division 420, without installing the display playback program in the personal computer beforehand when output data 400 are read with a Personal Digital Assistant, a personal computer, etc., the image data display function 411 carries out the image data 421 after composition, and the voice data regenerative function 412 carries out display playback of the voice data 422 after composition, respectively.

[0021]

When outputting output data 400 to a seal 430, the printout of the image data 421 after composition is carried out to a seal 430 with a seal printing machine. When using voice data here, voice data needs to be beforehand changed into the image data 421 after composition. That is, since only image data can deal with seal printing, voice data needs to be changed into image data. What is necessary is just to equip an image composition output unit with this speech synthesis output function with a means to be able to consider various gestalten as the approach, for example, to change into the bar code for voice playback the voice which is a well-known

technique. Since a voice playback bar code will also be printed with image data if seal printing is performed, voice will be reproduced if this bar code is read by the voice playback bar code reader of dedication. moreover -- if it is the voice which the user blew -- the voice -- receiving -- speech recognition processing -- carrying out -- alphabetic data -- changing -- further -- the alphabetic data -- " -- Japanese butterbur -- it considers as dashi" and is processed into image data. it was created -- " -- Japanese butterbur -- if dashi" is stored as image pattern data and it is compounded with photography image data, when it prints on a seal, it can also print a user's voice with image data. moreover, the voice pattern data beforehand stored in the image composition output unit with this speech synthesis output function are used -- if it becomes -- the alphabetic data of the voice pattern data -- storing -- it -- " -- Japanese butterbur -- it is possible to use as dashi" etc.

[0022]

Moreover, voice is able in the case of music or a background sound, to change into the image data of a musical score and to output, if the voice is MIDI data. Moreover, it matches with such music, the artist name and detailed information of music are stored, and you may output as image data in the case of printing. Thus, in using voice data, after changing voice data into alphabetic data or image data, it compounds to the image data 421 after composition. Then, the image data 421 after composition is set and printed [output and] to the seal printing machine 430.

[0023]

What is necessary is just to store output data 400 in the pocket storage 440 as it is, when outputting output data 400 to the pocket storages 440, such as an IC card and SmartMedia. When the image composition output unit with this speech synthesis output function has two or more kinds of storage output means, the class of storage is outputted after making a user specify one or plurality. Since it is stored in the pocket storage 440 while output data 400 have been digital data, if a user has a display regenerative apparatus with the usable pocket storage, he can reproduce output data 400 any number of times at any time. Even if the display playback program is not installed in the display regenerative apparatus by having had especially the program section 410, it is possible to carry out display playback of the data division 420. Furthermore, if the display regenerative apparatus is equipped with the image printing equipment, it can print additionally any number of times. Moreover, in an image composition output unit with this speech synthesis output function, a user can edit arbitration into the outputted data, such as adding and outputting processing and composition further, or newly creating output data as image pattern data or voice pattern data using a part of the data.

[0024]

When distributing output data 400 via a network, a user is made to specify a distribution place mail address, and output data 400 are distributed to the appointed address as it is. As it was sent to the mail address specified by a user while output data 400 had been digital data, and stated previously, of course, a user is able to print additionally any number of times, or to edit arbitration. Moreover, since network distribution is not influenced of a medium unlike the case of a pocket storage, if it is the environment where the terminal unit by the side of a user is equipped with the display playback program of output data, or the program section 410 can operate, a user can perform output data 400 for a display, playback, processing, edit, etc. with a terminal unit at any time. Furthermore, it is possible to transmit image data and voice data to an image composition output unit with this speech synthesis output function from a user's terminal side via a network. Thereby, the image composition output unit with this speech synthesis output function can be used as the image pattern data used in case the data by the side of a user are outputted to various media, such as a seal and an IC card, or overwrite processing of the data is carried out and new output data are created [**** / creating new output data], or voice pattern data.

[0025]

Next, it is based on the gestalt of some operations and explains.

Drawing 5 is a flow chart for explaining the processing at the time of applying an image composition output unit with this speech synthesis output function to "horoscopy." Hereafter,

an example of processing is explained, referring to drawing. First, a user is photoed with a camera 220 and let it be photography image data (step 501). Through the command input of a touch panel, such as "changing a color" etc., by processing photography image data by the dialogue, a user and an image composition output unit with this speech synthesis output function compound image pattern data to photography image data, and create a suitable photography image for a user (step 502). ["which compounds a background"] Next, an input is received for the date of birth by the user (step 503).

[0026]

The image composition output unit with this speech synthesis output function performs the program of horoscopy on the inputted date of birth, and asks for the language showing "today's fortune" (step 504). In addition, the result of horoscopy considers language which expresses "today's fortune" also with bad fortune even if as an affirmative expression. for example, -- "-- there is nothing the night not breaking. It will do its best to a slight degree. The soft expression " is used. next, the language showing "today's fortune" for which it asked -- Japanese butterbur -- the photography image which compounded to dashi image data and was created at step 502, and Japanese butterbur -- dashi image data is compounded and it considers as the image data after composition (step 505). here -- Japanese butterbur -- dashi image data -- Japanese butterbur -- it is image pattern data with which the language of "today's fortune" was written in dashi. Next, let the voice pattern data applicable to the language of "today's fortune" be the voice data of the language of ejection and "today's fortune" out of the voice data beforehand prepared for the image composition output unit with this speech synthesis output function (step 506). And a user is made to choose a background sound / background music (step 507). The sound pattern data of selected background sound / background music are compounded with ejection, it is compounded with the voice pattern data of the language of "today's fortune", and it considers as the voice data after composition (step 508). And a user is made to choose whether output data are outputted to a seal (step 509).

When the selection outputted by the user is made, the image data after composition is printed on a seal using a seal printing machine (step 510). Next, a user is made to choose whether output data are outputted to an IC card (step 511). When the selection outputted by the user is made, after [composition] image data, after [composition] voice data, and a display playback program are outputted to an IC card (step 512). A user is made to choose whether output data are distributed in a network finally (step 513). When the selection distributed by the user is made, a user is made to specify and do the mail address to distribute (step 514), and after [composition] image data, after [composition] voice data, and a display playback program are distributed to the mail address specified by [the] a user (step 515).

[0027]

Drawing 6 , drawing 7 , and drawing 8 are the examples of an output of the gestalt of operation of the application "horoscopy" in the image composition output unit with this speech synthesis output function shown in drawing 5 .

Drawing 6 is the example of 1 output which printed output data on the seal. A seal 600 is created based on the image data after composition. the image data after composition -- the photography image 601 and Japanese butterbur -- it is constituted by dashi 602. The photography image 601 is a photography image with which the user was obtained by leading command input actuation of a touch panel, and processing and compounding photography image data at an image composition output unit with this speech synthesis output function, and a dialogue, and the sun and the clouds in the background of drawing are also added by the user in the process. moreover, Japanese butterbur -- the language showing "today's fortune" from which dashi 602 was obtained by horoscopy -- Japanese butterbur -- it compounds as dashi image data.

Drawing 7 is the example of 1 output which outputted output data to the IC card. IC card 700 stores the image data 701 after composition, the voice data 702 after composition, and the display playback program 703. Here, the background sound / background music which the user itself chose in addition to the voice data applicable to the language of "today's fortune" are stored in the voice data 702 after composition.

Drawing 8 is the application of the above-mentioned example of an output, and sticks a seal 802 on IC card 801. When the content of IC card 801 looks at a seal 802 by this, it is effective in the ability to judge in an instant. Attachment of a seal 802 may be performed automatically by machine, may output IC card 801 and a seal 802 independently, and may have them stuck for user itself if needed for a user.

[0028]

Drawing 9 is a flow chart for explaining the processing at the time of applying an image composition output unit with this speech synthesis output function to "the message to a sweetheart/friend." Hereafter, an example of processing is explained with reference to drawing. A user is first photoed with a camera 220 and let it be photography image data (step 901). Through the command input of a user's touch panel, an image composition output unit with this speech synthesis output function processes photography image data by the dialogue, and creates a suitable photography image for a user (step 902). Next, a user receives "the message to a sweetheart/friend" blown from the microphone, and considers as sound recording voice data (step 903). The blown sound recording voice data is reproduced by the loudspeaker, the check of the content will be progressed to a user at a degree, if it is inquiry O.K., otherwise, the entrainment of a message is redone (step 904). next, the Japanese butterbur after the image composition output unit with this speech synthesis output function performs speech recognition processing of the message of the user who is sound recording voice data and changes a message into alphabetic data -- it compounds as dashi image data (step 905). and the photography image created at step 902 and Japanese butterbur -- dashi image data is compounded and let it be the image data after composition (step 906).

[0029]

Next, a user is made to choose a background sound / background music (step 907). Selected background sound / background music, and the message which is the sound recording voice data which the user blew are compounded, and it considers as the voice data after composition (step 908). And a user is made to choose whether output data are outputted to a seal printing machine (step 909).

[0030]

When the selection outputted by the user is made, the image data after composition is printed on a seal (step 910). Next, a user is made to choose whether output data are outputted to an IC card (step 911). When the selection outputted by the user is made, after [composition] image data, after [composition] voice data, and a display playback program are outputted to an IC card (step 912). A user is made to choose whether output data are distributed in a network finally (step 913). When the selection distributed by the user is made, a user is made to specify and do the mail address to distribute (step 914), and after [composition] image data, after [composition] voice data, and a display playback program are distributed to the mail address specified by [the] a user (step 915).

[0031]

Drawing 10 is the example of 1 output of the gestalt of operation of "the message to a sweetheart/friend" in the image composition output unit with this speech synthesis output function shown in drawing 9. In this example of an output, the seal 1000 is stuck on IC card 1001. It is the photography image which the user processed the photography image 1002 into the photography image data of a basis, and was obtained, and the star in the background of drawing is also added by the user in the process. moreover, Japanese butterbur -- the Japanese butterbur after changing into alphabetic data the message which the user blew for accumulating dashi 1003 in a sweetheart or a friend by speech recognition -- it compounds to dashi image data. IC card 1001 stores the image data after composition, the voice data after composition, and a display playback program.

Here, the background sound / background music which the user itself chose in addition to the message which the user blew are stored in the voice data after composition.

[0032]

Drawing 11 is a flow chart for explaining the processing at the time of applying an image composition output unit with this speech synthesis output function for "copying with a pet."

Hereafter, an example of processing is explained with reference to drawing. A user is first photoed with a camera 220 and let it be photography image data (step 1101). Through the command input of a user's touch panel, an image composition output unit with this speech synthesis output function processes photography image data by the dialogue, creates the suitable photography image for a user, and makes the photography image user image data (step 1102). Next, a user memorizes the message blown with the microphone as sound recording voice data (step 1103), and the background sound / background music which the user chose are taken out from voice pattern data (step 1104). This message, and a background/background music are compounded, and it considers as user voice data (step 1105). Next, the IC card which the user inserted is received (step 1106). And the image data after composition in an IC card and the voice data after composition are taken out, and it considers as pet image data and pet voice data, respectively (step 1107, step 1108). As for this IC card, a user photos his pet in an image composition output unit with this speech synthesis output function before photography. That is, a pet's image is stored in the image data after composition in an IC card, and a pet's voice is stored in the voice data after composition, respectively. Next, user image data and pet image data are compounded, and let it be the image data after composition (step 1109). Furthermore, user voice data and pet voice data are compounded, and let it be the voice data after composition (step 1110). And a user is made to choose whether output data are outputted to a seal printing machine (step 1111). When the selection outputted by the user is made, the image data after composition is outputted to a seal printing machine (step 1112). Next, a user is made to choose whether output data are outputted to an IC card (step 1113). When the selection outputted by the user is made, after [composition] image data, after [composition] voice data, and a display playback program are outputted to an IC card (step 1114). A user is made to choose whether output data are distributed in a network finally (step 1115). When the selection distributed by the user is made, a user is made to specify and do the mail address to distribute (step 1116), and after [composition] image data, after [composition] voice data, and a display playback program are distributed to the mail address specified by [the] a user (step 1117).

[0033]

Drawing 12 is the IC card which the user inserted in the image composition output unit with this speech synthesis output function in the gestalt of operation of the application "it copies with a pet" shown in drawing 11.

The after [composition] image data 1201, the after [composition] voice data 1202, and the display playback program 1203 are stored in IC card 1200. The image data 1201 after composition is the photography image of a pet's cat. The voice data 1202 after composition is the cry "MYAOMYAO" of the cat.

[0034]

Drawing 13 is an example of an output to the IC card in the gestalt of operation of the application "reflected with a pet" of the image composition output unit with this speech synthesis output function shown in drawing 11. A pet's data to be used are incorporated from the IC card shown in drawing 12. The after [composition] image data 1301, the after [composition] voice data 1302, and the display playback program 1303 are stored in IC card 1300. The image data 1301 after composition compounds the photography image of the cat of an IC card in a user's photography image. the voice data 1302 after composition -- a user's "next time play -- coming -- the bottom time -- " -- ** -- the cry "MYAOMYAO" of the cat incorporated from the IC card with the message to say is also stored.

[0035]

Although the data for composition are incorporated from the IC card with the gestalt of operation shown by drawing 11, drawing 12, and drawing 13, it cannot be overemphasized that other pocket storages may be used.

For example, the data photoed from the pocket storage in a digital camera can be read, and it can consider as the image data for compounding. On the contrary, it is also possible to read the image data and voice data in a pocket storage, and to carry out overwrite processing and composition on it.

[0036]

Drawing 14 is a flow chart for explaining the processing at the time of applying an image composition output unit with this speech synthesis output function to a "coma rate drama." A thing called karaoke makes Calah the part which a singer sings by music, and when a user compensates the part with his singing voice, it enables it to experience a little the temper from which the amateur became a singer generally. Although the coma rate drama described to it here is a still picture, it makes Calah the part which the heroine (or hero) performs in a drama, and can say that the temper from which the user became the heroine (or hero) of a drama enables it to taste a little by compensating the part with a user's portrait and voice. As a user's supporting actor, the character of a popular actor and animation etc. can be considered in that case.

[0037]

Hereafter, an example of processing of the flow chart of drawing 14 is explained, referring to the coma image transition diagram of drawing 15.

A user is made to choose a character to perform among the characters which first come out in a favorite coma rate drama and its drama out of two or more coma rate dramas prepared beforehand by the touch panel (step 1401). The number of coma of a coma rate drama may be decided beforehand, and before it starts photography, you may have a user specify it. When the number of coma increases, accounting is required according to the number of coma. And the coma of the beginning of the drama which the user chose is read (step 1402). Next, before displaying a coma, the character which the user chose from coma is eliminated, and the location where the character was displayed is displayed on a screen [else] (step 1403). A user's photography image is inserted in this location distinguished and displayed. (A) of drawing 15 is the coma image of a basis, and (B) is the processed coma image of step 1403. 1501 of (A) is the user selection character which the user chose, and it inserts in a user image instead of being this character image. In order to eliminate the user selection character 1501 by processing of step 1403 and to display on the location [else], for example, as shown in drawing, half tone dot meshing 1502 is performed. A user is inserted in this half-tone-dot-meshing part 1502. A favorite pose is made looking at the processed coma image (B) with which the user was displayed, a photograph is taken with a camera, and it considers as photography image data (step 1404). A user rotates a revolution base, or operates the upper and lower sides of a camera, and may enable it to take a photograph by the favorite angle type at this time. And photography image data is compounded with the processed coma image (B) of drawing 15 (step 1405). It is the after [composition] coma image in which the result was shown by (C) of drawing 15. A user's image 1503 is inserted in the half-tone-dot-meshing part 1502 instead of a user selection character, and it is displayed on it. Next, the conversation of a character is directed to a user. First, the words of the character assigned to a user of a coma under current photography are displayed (step 1406). In the example of drawing 15, the words "whether it will clear up tomorrow" will be displayed on a screen. And a user blows words according to the display, and stores the voice data (step 1407). And the voice data of each character required for the conversation within a coma is read, and the voice data is stored (step 1408). Incidentally the conversation of each character and the voice data of words are recorded beforehand. For example, the words of the supporting actor character of drawing 15 of "thinking that it probably clears up" are beforehand recorded as voice pattern data. Furthermore, the playback sequence of the words of each character is the array which a natural conversation hears, when it reproduces. That is, since it is stored instead of the words which the user blew being voice pattern data of a user selection character, when voice is reproduced, it is audible as the user is talking with other characters. The voice data and BGM of these conversation, and a sound effect are compounded, and there is coma voice after composition (step 1409). Finally, finally the coma image 1503 after composition is displayed on a screen, voice playback of all the conversation within a coma including the words which the user blew is carried out, and it checks to a user whether this is sufficient (step 1410). If not pleased, return, photography, and sound recording are redone to step 1404.

Since the creation for one coma finished above, it judges whether the coma of a continuation is in a degree (step 1411). If there is a coma of a continuation, the following coma will be read and it will return to step 1403 (step 1412). If there is no coma of a continuation and creation of all

coma finishes, it will go into output processing. The after [composition] coma image for all coma and the coma voice after composition are after [composition] image data, and after [composition] voice data, respectively. First, a user is made to choose whether output data are printed on a seal (step 1413). When the selection outputted by the user is made, the image data after composition is printed (step 1414).

. Next, a user is made to choose whether output data are outputted to an IC card (step 1415). When the selection outputted by the user is made, after [composition] image data, after [composition] voice data, and a display playback program are outputted to an IC card (step 1416). A user is made to choose whether output data are distributed in a network finally (step 1417). When the selection distributed by the user is made, a user is made to specify the mail address to distribute (step 1418), and after [composition] image data, after [composition] voice data, and a display playback program are distributed to the mail address specified by [the] a user (step 1419).

[0038]

Drawing 16 is an example of the output to an IC card in the gestalt of operation of the application "a coma rate drama" of an image composition output unit with this speech synthesis output function. The image data 1601 after composition, the voice data 1602 after composition, and the display playback program 1603 are stored in IC card 1600. The image data 1601 after composition consists of two or more coma images after composition, and consists of four coma in this case. And the woman shown in drawing is a user entrapped instead of the user selection character. The conversation of a coma rate drama is stored in the voice data 1602 after composition as voice. The part by which the underline is drawn is the part which the user took charge of, i.e., the words which the user blew from the microphone. And although 1603 is a display playback program, it cannot stop at only displaying an image as the display playback program in the gestalt of this operation, and reproducing voice, but the coma voice after composition can be reproduced corresponding to each coma which displayed and displayed the coma image after composition in order of a coma.

[0039]

Although it is clear that this design can reproduce not only image data but voice data according to the gestalt of operation of above this design, the playback approach of the voice by this design does not stop at the gestalt of operation of this design. The example of the playback approach considered is explained with reference to drawing the following. by the way -- the image data after composition currently printed since voice data is reproduced in the following examples of an output with voice -- "Japanese butterbur -- dashi" does not use.

[0040]

Drawing 17 is an example which performed voice playback only by seal printing. The bar code 1702 for voice playback other than the printing image 1701 is printed by the seal 1700. This uses the technique which changes into the bar code for voice playback the voice which is a well-known technique, and is printed. What is necessary is just to read the bar code 1702 for voice playback printed by the seal 1700 by the bar code reader of the dedication which is not illustrated, when reproducing voice.

[0041]

Drawing 18 is an example of the playback approach which carries out voice playback only with a pocket storage. Since the printed seal 1801 is stuck on IC card 1800, image data is obtained by seeing a seal 1801. About voice data, it is reproducible by operating the voice playback component 1802 in IC card 1800. That is, an image and voice can be enjoyed only with IC card 1800.

[0042]

Drawing 19 is an example of the playback approach reproduced in voice data with a personal digital assistant. Key case mold IC card READER1900 is a small personal digital assistant, and is equipped with an image display means and a voice playback means. By this, a display playback program is started, the image data after composition is displayed on a display 1901, and the voice data after composition is outputted from the voice output section 1902.

[0043]

Drawing 20 is an example of the playback pattern which carries out display playback of the image data after composition, and the voice data after composition with a personal computer 2000. It read from IC card READER2001, or the output data (after [after / composition / image data + composition] voice data + display playback program) which arrived as e-mail in the network 2002 course are received, and a personal computer 2000 reads and starts a display playback program, and is displayed and reproduced by the monitor 2003 and the built-in loudspeaker 2004.

[0044]

As other playback patterns, although not illustrated, there is a playback pattern using the image composition output unit with this speech synthesis output function itself. For that purpose, what is necessary is just to give the function which reads output data into the image composition output unit with this speech synthesis output function itself from an IC card, and is displayed and reproduced.

[0045]

[Effect of the Device]

As explained above, this design can provide a user with the amusement environment more near reality which added voice to the image by having added the function in which sound data, such as background sounds, such as voice of not only image data but the recorded user and a natural sound, and background music, can be treated to the conventional image printing equipment. Moreover, by adding a display playback program to image data and sound data, and performing distribution to the writing to a pocket storage, and the assignment mail address via a network, a user can display image data and sound data and can be reincarnated later.

[Translation done.]

*** NOTICES ***

JPO and NCIPi are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.**** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

[Drawing 1] It is drawing showing the hardware configuration of the image composition output unit with a speech synthesis output function of this design.

[Drawing 2] It is the block diagram showing the important section of the configuration of 1 operation gestalt of an image composition output unit with this speech synthesis output function.

[Drawing 3] It is the outline perspective view showing the concrete configuration in the case of performing an information storage medium output.

[Drawing 4] It is drawing showing the structure and the output destination change of output data.

[Drawing 5] It is the flow chart which shows the processing at the time of applying an image composition output unit with this speech synthesis output function to "horoscopy."

[Drawing 6] It is the example of 1 output which printed output data on the seal with the gestalt of operation of "horoscopy."

[Drawing 7] It is the example of 1 output which outputted output data to the IC card with the gestalt of operation of "horoscopy."

[Drawing 8] It is the example of an output which stuck the seal on the IC card with the gestalt of operation of "horoscopy."

[Drawing 9] It is the flow chart which shows the processing at the time of applying an image composition output unit with this speech synthesis output function to "the message to a sweetheart/friend."

[Drawing 10] It is an example of 1 output in the gestalt of operation of "the message to a sweetheart/friend."

[Drawing 11] It is the flow chart which shows the processing at the time of applying an image composition output unit with this speech synthesis output function for "copying with a pet."

[Drawing 12] In the gestalt of operation of "copying with a pet", it is the explanatory view in which a user shows the IC card inserted in an image composition output unit with this speech synthesis output function.

[Drawing 13] It is drawing showing the example of 1 output in the gestalt of operation of "copying with a pet."

[Drawing 14] It is the flow chart which shows the processing at the time of applying an image composition output unit with this speech synthesis output function to a "coma rate drama."

[Drawing 15] It is a coma image transition diagram in the gestalt of operation of a "coma rate drama."

[Drawing 16] It is drawing showing the example of 1 output in the gestalt of operation of a "coma rate drama."

[Drawing 17] It is drawing showing an example which performed voice playback only by seal printing.

[Drawing 18] It is the explanatory view showing an example which performed voice playback only with the pocket storage.

[Drawing 19] It is the explanatory view showing an example which performed voice playback with

the personal digital assistant.

[Drawing 20] It is the explanatory view showing an example which performed voice playback with the personal computer.

[Explanation of agreement]

1 [-- Revolution base,] -- A monitor-cum-a touch panel, 2 -- A microphone, 3 -- A loudspeaker, 4 5 [-- A processor, 9 / -- Network,] -- A driving gear, 6 -- A camera, 7 -- A driving gear, 8 10 -- A seal printing machine, 11 -- READ/WRITE equipment of media, such as an IC card, 400 -- Output data, 410 -- The program section, 411 -- Image data display function, 412 [-- After / composition / voice data, 430 / -- A seal, 440 / -- Pocket storages, such as an IC card, 450 / -- Mail address specified by a user.] -- A voice data regenerative function, 420 -- Data division, 421 -- After [composition] image data, 422

[Translation done.]

(19) 日本国特許庁 (J P)

(12) 登録実用新案公報 (U)

(11) 実用新案登録番号

第3060617号

(45) 発行日 平成11年(1999) 9月 7 日

(24) 登録日 平成11年(1999) 6月16日

(51) Int.Cl. ⁸	識別記号	F I		
H 0 4 N	5/765	H 0 4 N	5/781	5 1 0 H
	5/781	G 0 9 F	27/00	N
G 0 9 F	27/00	G 0 9 G	5/00	5 1 0 Q
G 0 9 G	5/00	G 1 0 L	3/00	S
G 1 0 L	3/00		5/02	J
評価書の請求 未請求 請求項の数 4 O L (全 30 頁) 続き有				

(21) 出願番号 実願平10-10405

(22) 出願日 平成10年(1998)12月28日

(73) 実用新案権者 000233055

日立ソフトウェアエンジニアリング株式会
社

神奈川県横浜市中区尾上町 6 丁目 81 番地

(72) 考案者 井上 明

神奈川県横浜市中区尾上町 6 丁目 81 番地
日立ソフトウェアエンジニアリング株式会
社内

(74) 代理人 弁理士 秋田 収喜

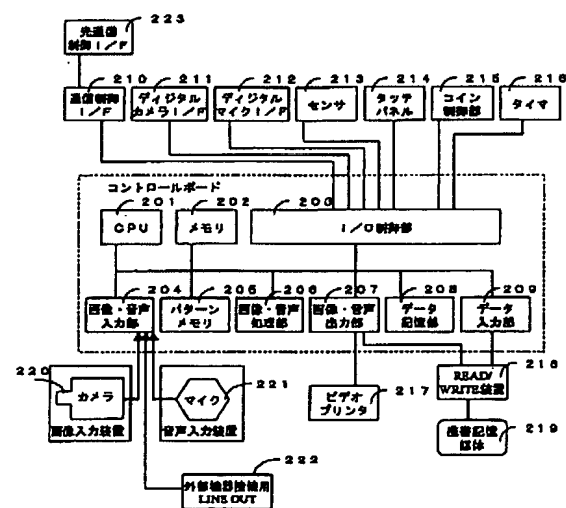
(54) 【考案の名称】 音声合成出力機能付画像合成出力装置

(57) 【要約】

【課題】 画像と音声の両方を取り扱えるようにして、目と耳の両方からユーザに臨場感を与えることができる音声合成出力機能付画像合成出力装置を提供すること。画像データ及び音声データの表示・再生の他に、自由な編集・加工を可能にすること。

【解決手段】 ユーザの声を録音しメモリに記憶させる録音手段と、背景音、背景音楽等の複数の音データを記憶した音データ記憶手段と、この音データ記憶手段に記憶された複数の音データのうちユーザが選択した音データと前記録音手段によって録音したユーザの音声データとを合成して出力する音合成出力を設ける。また、合成した画像データ及び音データを各種記憶媒体に出力、またはネットワーク経由で指定メールアドレスへ配信し、後で各種記憶媒体に出力された画像データ及び音データを入力し、モニタに表示、スピーカから再生可能に構成する。

図 2



1

【実用新案登録請求の範囲】

【請求項1】 被写体を撮影する撮影手段と、
被写体が発生した音声を録音する録音手段と、
前記撮影手段によって撮影した撮影画像データを記憶する
撮影画像データ記憶手段と、
前記録音手段によって録音した録音音声データを記憶する
録音音声データ記憶手段と、
前記撮影画像データと合成する複数の画像パターンデータ
を記憶した画像パターンデータ記憶手段と、
前記録音音声データと合成する複数の音パターンデータ 10
を記憶した音パターンデータ記憶手段と、
複数の画像パターンデータのうち所望の画像パターンデータ
を選択する画像パターンデータ選択手段と、
前記撮影画像データに前記画像パターンデータ選択手段
で選択された画像パターンデータを合成して合成後画像
データを出力する画像合成出力手段と、
複数の音パターンデータのうち所望の音パターンデータ
を選択する音声パターンデータ選択手段と、
前記録音音声データに前記音パターンデータ選択手段で 20
選択された音パターンデータを合成して合成後音声データ
を出力する音声合成出力手段とを備え、
さらに、前記合成後画像データをシールに印刷出力する
印刷手段、前記合成後画像データと前記合成後音声データ
と前記合成後画像データを表示し前記合成後音声データ
を再生するための表示再生プログラムとを携帯記憶媒体
へ格納する格納手段、前記合成後画像データと前記合成
後音声データと前記合成後画像データを表示し前記合成
後音声データを再生するための表示再生プログラムとを
ネットワーク経由で利用者指定のメールアドレスに配信
する配信手段、のいずれか1つあるいは複数の備えた 30
音声合成出力機能付画像合成出力装置。
【請求項2】 前記録音音声データを文字データに変換し、
その変換された文字データを画像パターンデータに合成して、
前記画像パターンデータ記憶手段に格納する音声データ
変換手段をさらに備えることを特徴とする請求項1記載の
音声合成出力機能付画像合成出力装置。
【請求項3】 前記携帯記憶媒体内に予め記憶された画像
データまたは音データあるいは両方を読取り、画像データ
を前記撮影画像データか画像パターンデータとして、
音データを録音音声データか音パターンデータとして、 40
それぞれ入力する携帯記憶媒体読取入力手段をさらに備える
ことを特徴とする請求項1記載の音声合成出力機能付画像
合成出力装置。
【請求項4】 ネットワーク経由の画像データまたは音
データあるいは両方を受信し、受信した画像データを前
記撮影画像データか画像パターンデータとして、音データ
は録音音声データか音パターンデータとして、それぞれ
入力するデータ受信入力手段をさらに備えることを特徴
とする請求項1記載の音声合成出力機能付画像合成出力
装置。

2

【図面の簡単な説明】

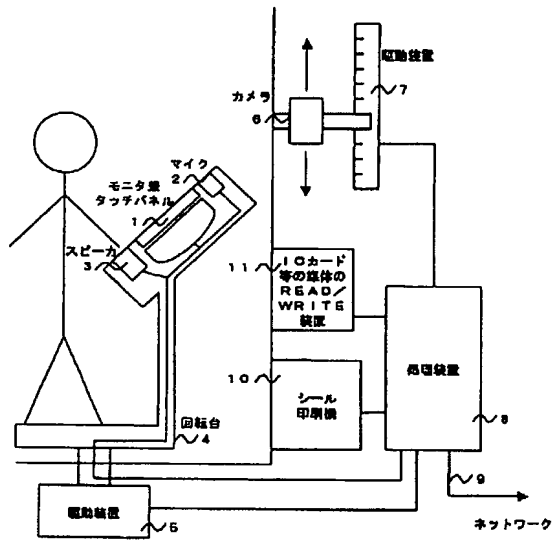
【図1】 本考案の音声合成出力機能付画像合成出力装置
のハードウェア構成を示す図である。
【図2】 本音声合成出力機能付画像合成出力装置の一実
施形態の構成の要部を示すブロック図である。
【図3】 情報記憶媒体出力を行なう場合の具体的な構成
を示す概略斜視図である。
【図4】 出力データの構造と出力先を示す図である。
【図5】 本音声合成出力機能付画像合成出力装置を「星
占い」に適用した場合の処理を示すフローチャートである。
【図6】 「星占い」の実施の形態で、出力データをシール
に印刷した一出力例である。
【図7】 「星占い」の実施の形態で、出力データをIC
カードに出力した一出力例である。
【図8】 「星占い」の実施の形態で、シールをICカード
に貼り付けた出力例である。
【図9】 本音声合成出力機能付画像合成出力装置を「恋人
／友人へのメッセージ」に適用した場合の処理を示す
フローチャートである。
【図10】 「恋人／友人へのメッセージ」の実施の形態
における一出力例である。
【図11】 本音声合成出力機能付画像合成出力装置を
「ベットと写す」に適用した場合の処理を示すフロー
チャートである。
【図12】 「ベットと写す」の実施の形態において、ユー
ザが本音声合成出力機能付画像合成出力装置に挿入した
ICカードを示す説明図である。
【図13】 「ベットと写す」の実施の形態における一出
力例を示す図である。
【図14】 本音声合成出力機能付画像合成出力装置を
「コマ割りドラマ」に適用した場合の処理を示すフロー
チャートである。
【図15】 「コマ割りドラマ」の実施の形態におけるコマ
画像遷移図である。
【図16】 「コマ割りドラマ」の実施の形態における一出
力例を示す図である。
【図17】 シール印刷のみで音声再生を行なった一例を
示す図である。
【図18】 携帯記憶媒体のみで音声再生を行なった一例
を示す説明図である。
【図19】 携帯端末で音声再生を行なった一例を示す説
明図である。
【図20】 パソコンで音声再生を行なった一例を示す説
明図である。
【符合の説明】
1…モニタ兼タッチパネル、2…マイク、3…スピー
カ、4…回転台、5…駆動装置、6…カメラ、7…駆動
装置、8…処理装置、9…ネットワーク、10…シール
印刷機、11…ICカード等の媒体のREAD／WRITE

TE装置、400…出力データ、410…プログラム部、411…画像データ表示機能、412…音声データ再生機能、420…データ部、421…合成後画像データ

*タ、422…合成後音声データ、430…シール、440…ICカード等携帯記憶媒体、450…ユーザ指定のメールアドレス。

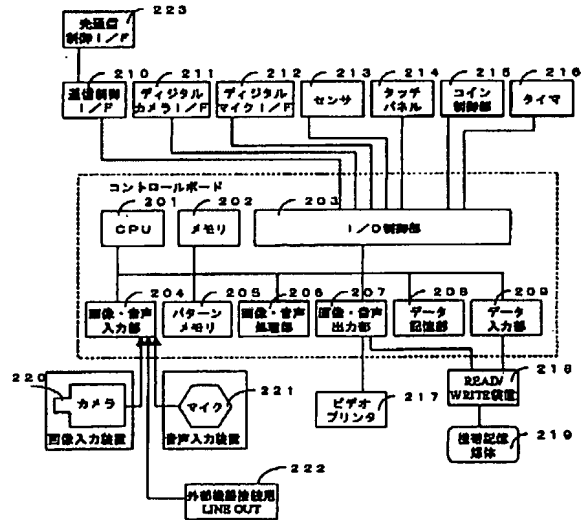
【図1】

図 1



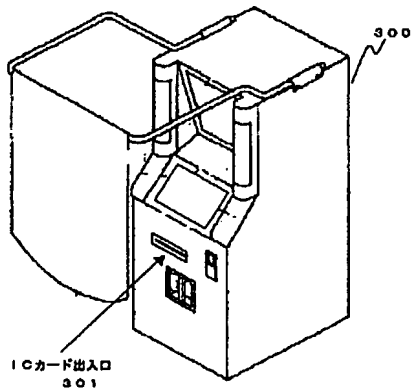
【図2】

図 2



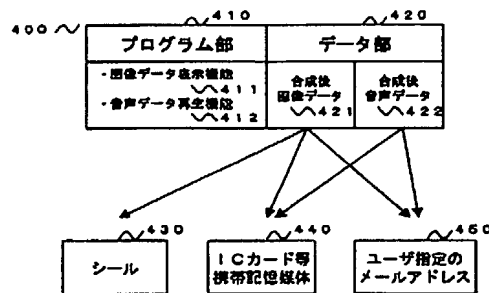
【図3】

図 3



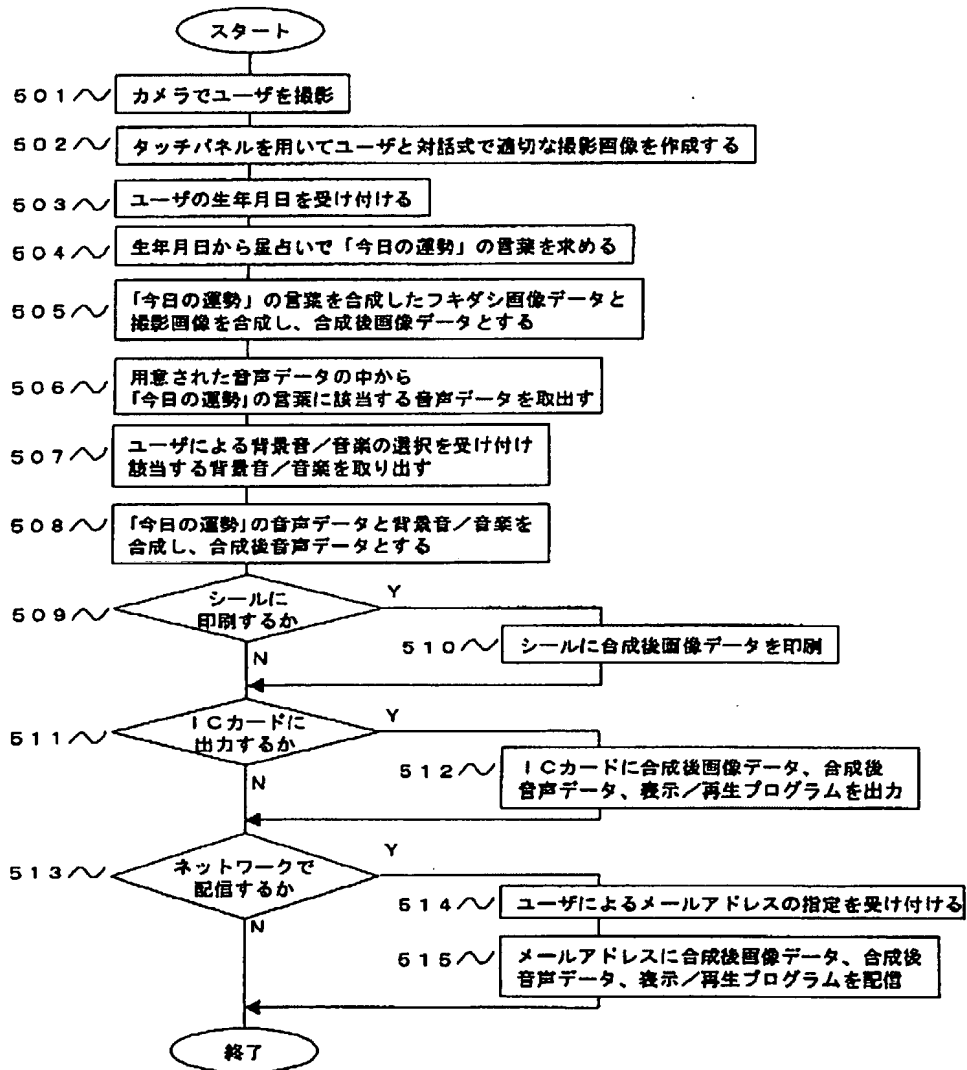
【図4】

図 4

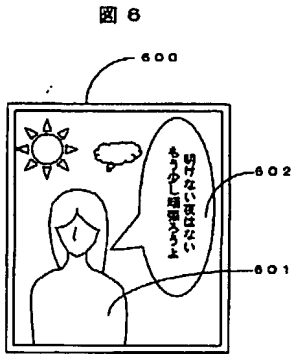


【図5】

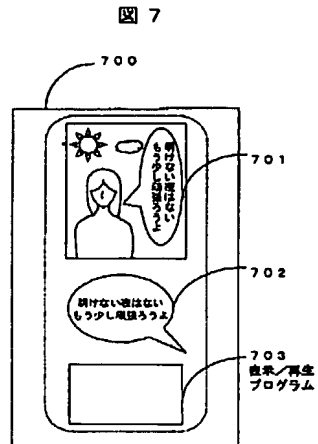
図 5



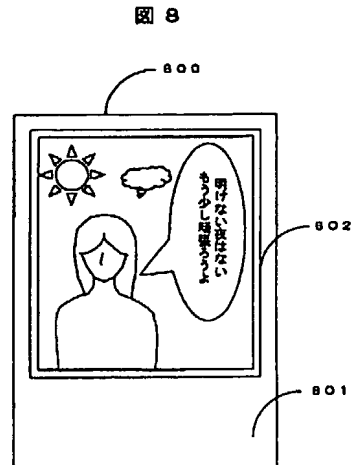
【図6】



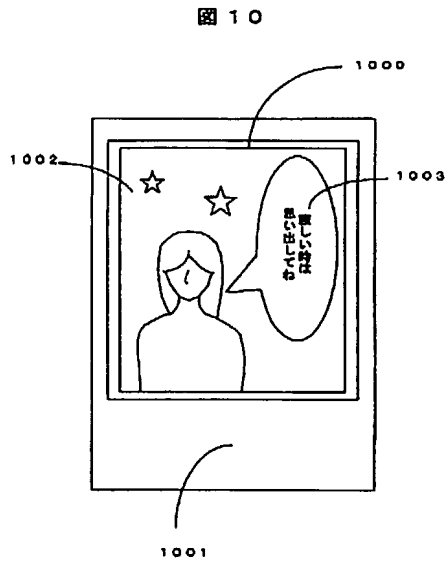
【図7】



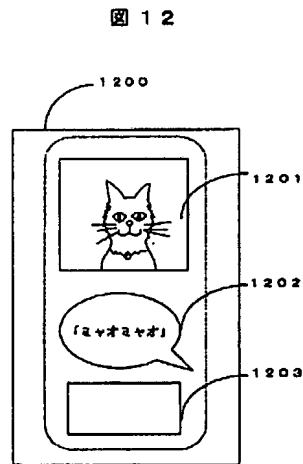
【図8】



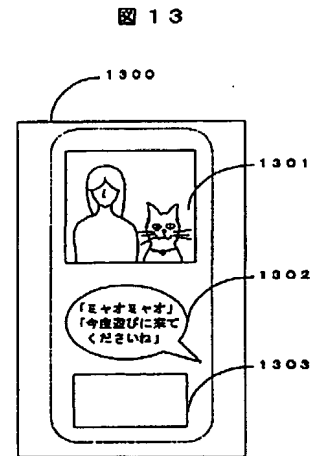
【図10】



【図12】

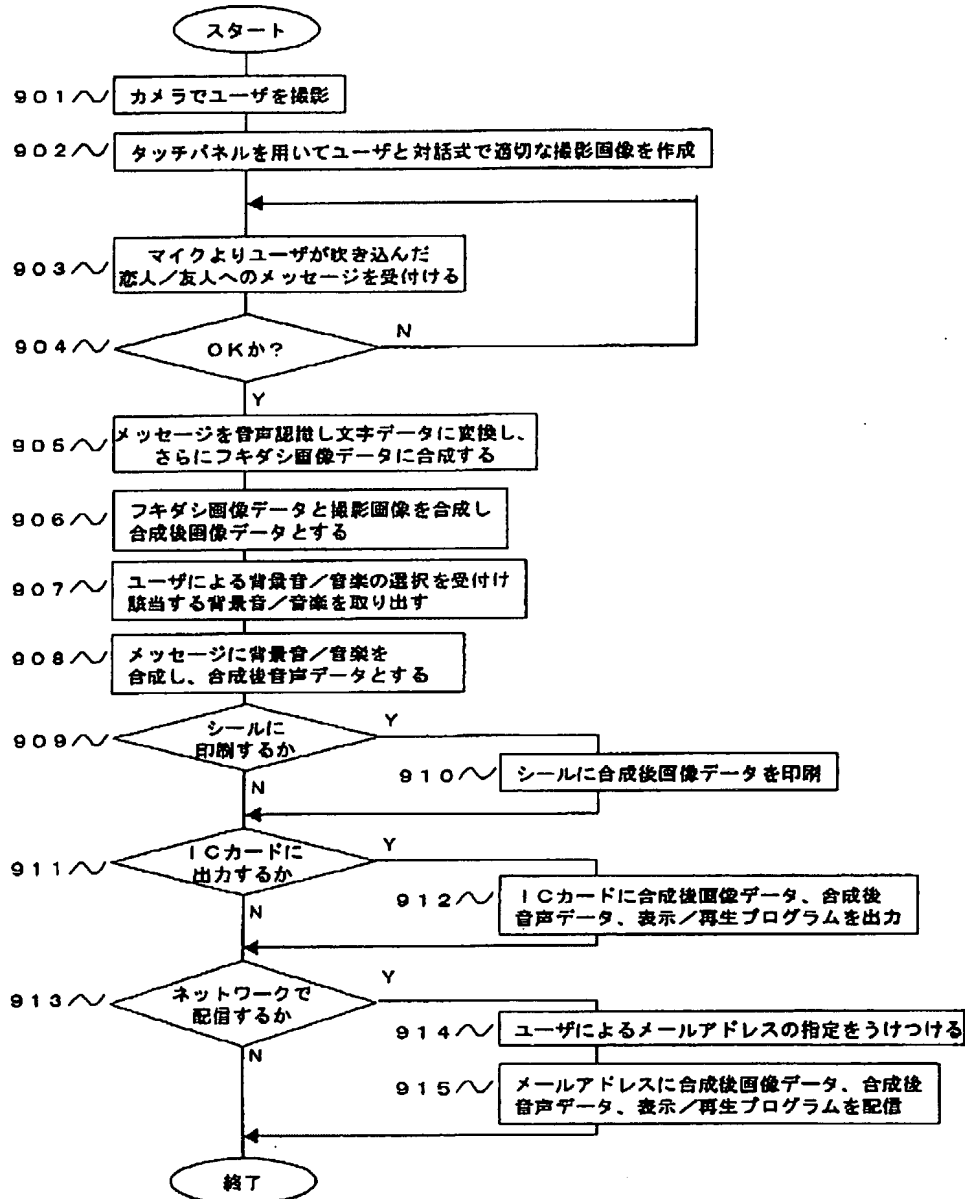


【図13】



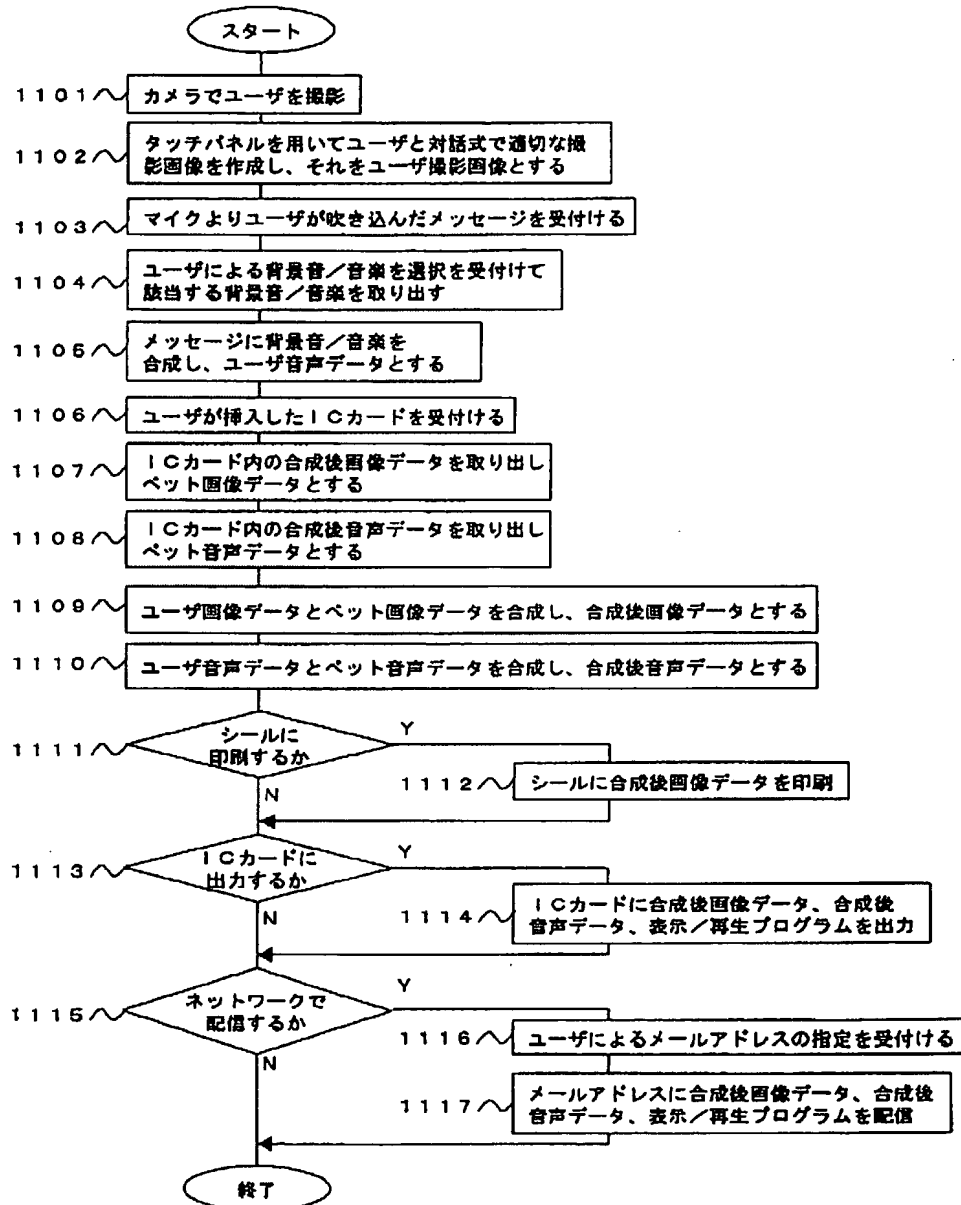
【図9】

図 9



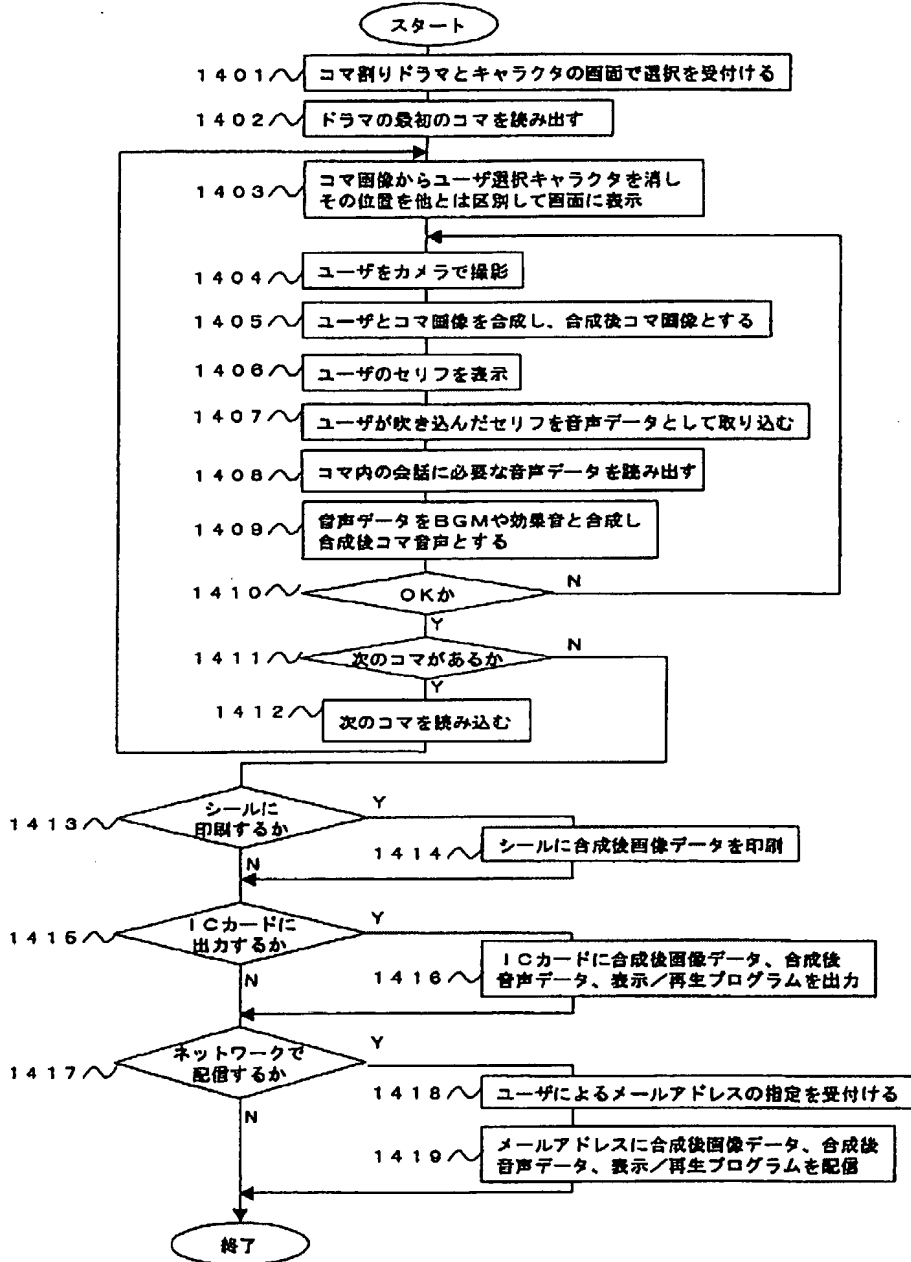
【図11】

図 11



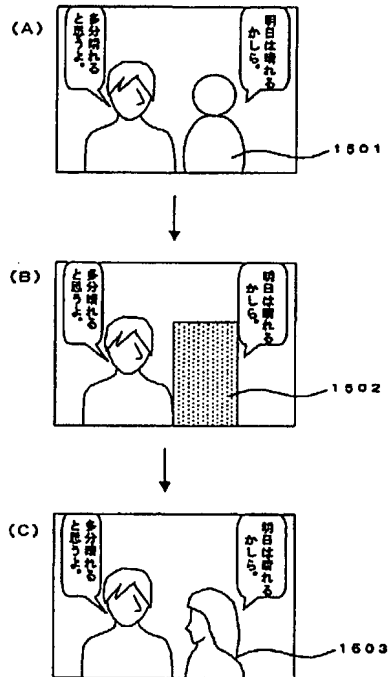
【図14】

図 14



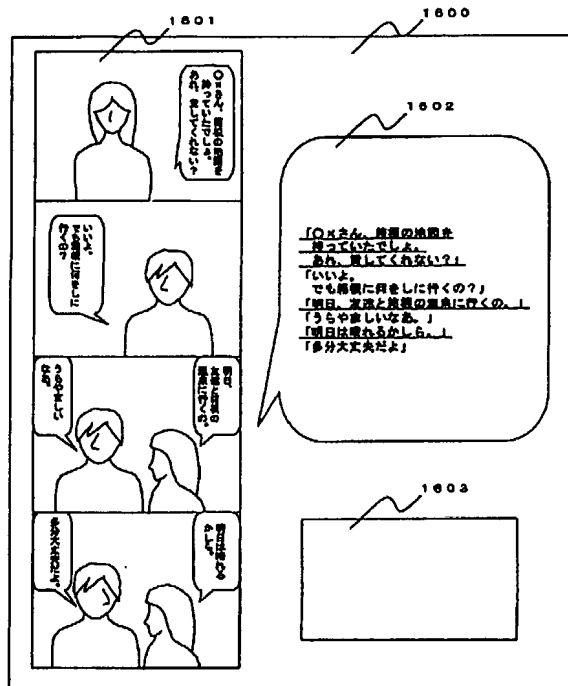
【図15】

図 15



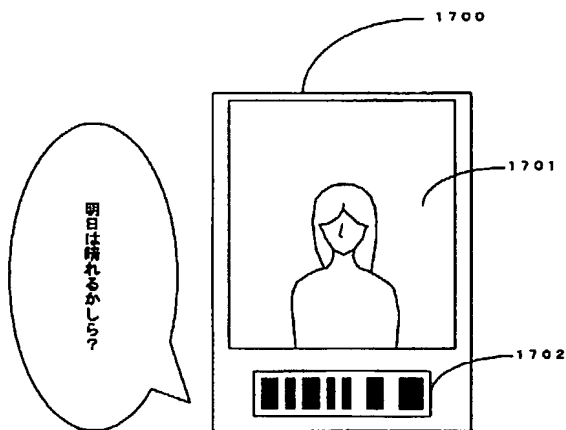
【図16】

図 16



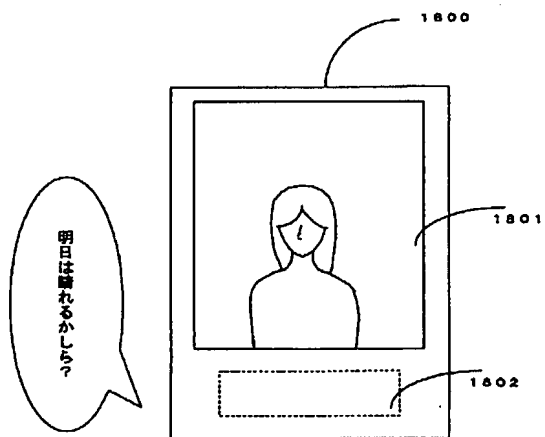
【図17】

図 17



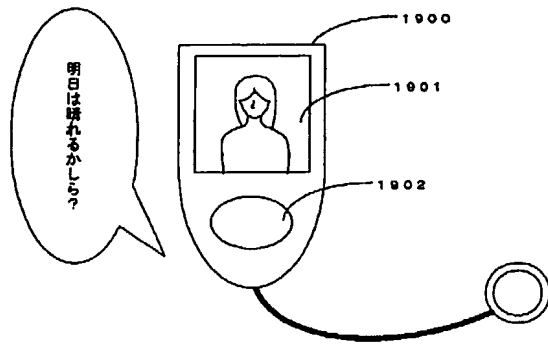
【図18】

図 18



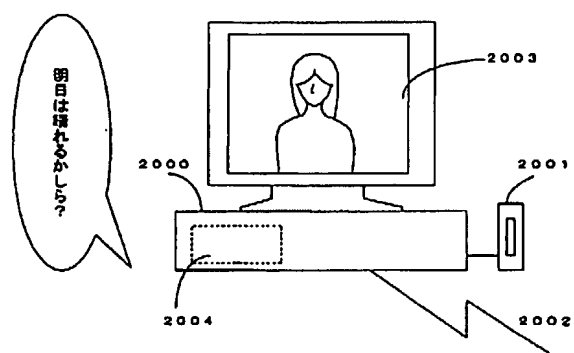
【図19】

図19



【図20】

図20



フロントページの続き

(51)Int.Cl.⁸
G10L 5/02
H04N 5/225

識別記号

F I
H04N 5/225 Z
5/781 510F

【考案の詳細な説明】

【0001】

【考案の属する技術分野】

本考案は、アミューズメントパークやゲームセンタ等に設置され、被写体(利用者)とあらかじめ用意された画像パターンとを合成して撮影し、その撮影画像をシールや名刺等としてプリントアウトする装置に係り、詳しくは、撮影映像のみでなく音声等の音も合成して所定の出力媒体から出力可能にした音声合成出力機能付画像合成出力装置に関するものである。

【0002】

【従来の技術】

近年、ユーザの肖像画像を撮影し、あらかじめ用意された背景等の画像データをユーザが指定することで、撮影された肖像画像と指定された画像データとを合成し、合成結果をプリントシート/シールや名刺などへ印刷する画像プリント装置がアミューズメントパークやゲームセンタ等に設置されている。

この画像プリント装置は、被写体であるユーザを撮影するためのカメラ、背景などの複数の合成用画像データをあらかじめ記憶したメモリ、撮影画像と背景等の合成用画像データを合成する画像合成手段、撮影画像や合成用画像データを表示するモニタ、合成画像をシールや名刺等に印刷するプリンタなどを基本的な構成として備えている。

【0003】

【考案が解決しようとする課題】

ところで、上記従来の画像プリント装置が画像データしか取り扱っていないことは、作成した合成画像をシールや名刺等に印刷するという上記の目的、そしてその目的を達成するための上記の構成から明らかである。従来の画像プリント装置では、カメラやモニタ、あるいは画像合成手段やプリンタなど、画像を取り扱うための手段のみで構成されており、音声や音楽などを装置内に取り込む集音、録音機構やそれら音声に関する処理を行なう手段は設けられていなかった。つまり、従来の画像プリント装置は音声や音楽、効果音など、音に関する情報を利用対象にしていなかったために、任意に録音されたユーザの声や自然音等の背景音

、背景音楽等の音声データを画像と一緒に取扱えないという問題があった。

【0004】

また、従来の画像プリント装置では、出力はプリントシート／シールや絵葉書などに印刷するという2次元的展開のみであり、合成画像の印刷終了後に、さらにその合成画像を編集・加工したり、あるいは焼き増しのように合成画像を再印刷するといったことはできない。また、従来の画像プリント装置は、ネットワークに接続されていても、背景パターンの書き換えや装置の保守に利用するだけであった。

【0005】

本考案の目的は、従来の画像プリント装置に、録音されたユーザの声、自然音等の背景音、背景音楽等の音声データを扱える機能を加えることにより、画像と音声の両方を取り扱えるようにして、目と耳の両方からユーザに臨場感を与えることができる音声合成出力機能付画像合成出力装置を提供することにある。

【0006】

また、本考案の他の目的は、画像や音声をデジタルデータとして取り扱い、出力手段としてプリントシート／シールへの印刷手段に加え、各種情報記憶媒体への格納手段、ネットワークを経由した指定メールアドレスへの配信手段を備えることにより、後に画像データ及び音声データの表示・再生を行えるようにし、またユーザがその画像なり音声に自由に編集・加工することも可能となるような音声合成出力機能付画像合成出力装置を提供することにある。

【0007】

【課題を解決するための手段】

本考案の目的は、ユーザの声を録音しメモリに記憶させる録音手段と、背景音、背景音楽等の複数の音データを記憶した音データ記憶手段と、この音データ記憶手段に記憶された複数の音データのうちユーザが選択した音データと前記録音手段によって録音したユーザの音声データとを合成して出力する音合成出力を設けることで達成される。

また、本考案の他の目的は、ICカード、フロッピーディスク等の各種記憶媒体への出力手段、ネットワーク経由の指定メールアドレスへの配信手段、各種記

憶媒体からの入力手段、画像データのモニタへの表示手段、音声データのスピーカからの再生手段を用いて、合成した画像データ及び音データを各種記憶媒体に出力、またはネットワーク経由で指定メールアドレスへ配信し、後で各種記憶媒体に出力された画像データ及び音データを入力し、モニタに表示、スピーカから再生することにより達成される。

【0008】

【考案の実施の形態】

以下、図面を参照して本考案の実施の形態を詳細に説明する。

図1は、本考案の音声合成出力機能付画像合成出力装置の一実施形態を示すハードウェア構成図である。

この実施形態の音声合成出力機能付画像合成出力装置には、モニタ兼タッチパネル1、マイク2、スピーカ3が設けられ、これらのモニタ兼タッチパネル1、マイク2、スピーカ3はユーザ(被写体)を乗せる回転台4に固定されており、回転台4が回転してもユーザはモニタを見てタッチパネル1を操作し、自分の声を録音し、またガイダンスや音声データの再生音を聞くことができるようになっている。また、回転台4は駆動装置5により回転し、カメラ6は駆動装置7により上下運動するように構成され、ユーザを自由な構図で撮影することができるようになっている。この場合、カメラ6の視野は、ユーザの頭部から足元までの全身画像を撮影可能に設定されている。ユーザから見えない装置内部には、画像処理等を行う処理装置8が設けられている。この処理装置8は、ネットワーク9に接続され、作成された画像データと音声データをネットワーク9を経由してユーザの指定したメールアドレスに配信することができるようになっている。処理装置8において加工、合成された画像データや音声データは、シール印刷機10においてシールに印刷出力される。また、ICカード等の媒体のREAD/WRITE装置11において、画像データや音声データをICカードのような携帯記憶媒体に出力することもできる。

【0009】

図2は、本考案の一実施形態の構成の要部を示すブロック図である。

本考案にかかる音声合成出力機能付画像合成出力装置のCPU201は、メモ

り202、I/O制御部203、画像・音声入力部204、パターンメモリ205、画像・音声処理部206、画像・音声出力部207、データ記憶部208、データ入力部209に接続されている。各回路ブロックは、例えば、公知の半導体集積回路応用技術によって半導体基板上、またはガラスエポキシ基板上に形成されるものである。

CPU201はデジタルカメラI/F211とマイクI/F212からの圧縮デジタル信号を受信し、その圧縮デジタル信号を画像・音声処理部206に格納されているプログラムにより合成し、ビデオプリンタ217(図1のシール印刷機10に相当)によるプリント出力処理やREAD/WRITE装置(図1のICカードのREAD/WRITE装置11に相当)218による情報記憶媒体への出力を行なう。

【0010】

メモリ202は画像・音声入力部204から取り込んだ画像データと音声データを記憶するほか、CPU201に発する命令と本音声合成出力機能付画像合成出力装置の制御手順となる命令群をプログラムとして記憶している。また、CPU201の処理データもワークメモリとして一時的に記憶する。

画像・音声入力部204は、カメラ(図1のカメラ6に相当)220が撮影した画像データとマイク(図1のマイク2に相当)221が録音した音声データをA/D変換装置(図示しない)によりアナログ信号からデジタル信号に変換する。このとき、画像撮影についてはクロマキー処理などを行なうことも可能である。

パターンメモリ205は、メモリ202にある画像・音声データと重ね合わせるための、背景やキャラクタ等の画像パターンデータと音楽やキャラクタの音声等の音パターンデータを記憶する。後述するように外部のデータを格納し、画像パターンデータや音声パターンデータとして利用できることはもちろん、メモリ202に格納された合成データそのものを、画像パターンデータや音パターンデータとして格納することも可能である。

【0011】

画像・音声処理部206は、メモリ202の画像・音声データとパターンメモリ205の画像パターンデータや音パターンデータをもとに、画像・音声の合成

処理を行なう。

画像・音声出力部207は、画像・音声処理部206が加工したデータを出力する。画像・音声出力部207はビデオプリンタ217とREAD/WRITE装置218に接続しており、ユーザの指示あるいは出力の形態に従って、ビデオプリンタ217やREAD/WRITE装置218に出力する。

データ記憶部208は、パターンメモリ205の内容を書き換えるものである。パターンメモリ205に書き込まれているデータの日付と照合し、データ記憶部208に記憶されたデータが新しいものであれば、パターンメモリ205内のデータを書き換える。これにより、データ記憶部208に記憶されたデータを背景、音声、キャラクタ等の画像パターンデータや音パターンデータとして用いる事ができる。

【0012】

データ入力部209はREAD/WRITE装置218に接続し、携帯記憶媒体219の画像データや音データなどを、デジタル情報のまま取り込むものである。データ入力部209は、取り込んだデータを一旦データ記憶部208に格納する。データ記憶部208は、そのデータをパターンメモリ205に書き込まれたデータの日付と比較し、パターンメモリ205のデータよりも新しいものと判断したら、パターンメモリ205内のデータをデータ記憶部208のデータに書換える。これにより、データ入力部209が取り込んだデータを、背景、音声、キャラクタ等の画像パターンデータや音パターンデータとして用いる事ができる。さらに、データ入力部209は取り込んだデータをメモリ202に格納し、そのデータに上書きして合成することも可能である。

【0013】

ところで、I/O制御部203は、通信制御I/F210、デジタルカメラI/F211、マイクI/F212、センサ213、タッチパネル214、コイン制御部215、タイマ216に接続し、これら全体の制御を行うものである。

通信制御I/F210は、外部ネットワークとの接続に用いる。通信制御I/F210は専用回線を介して、外部にある機器管理サーバに接続する。この機器管理サーバ（図示しない）は故障の有無を管理することを目的とする。また、新

しい背景パターン、音パターン、キャラクタ等を機器管理サーバから送信し、I/O制御部203を介してデータ記憶部208に記憶する。また通信制御I/F210は、光通信制御I/F223に接続している。光通信制御I/F223は赤外線通信を行なうIr-D Aの制御や、光ファイバーを利用した通信の制御を行なう。このインターフェースを用いることでプログラムなどのバイナリデータのやり取りができ、PDAなどの情報携帯端末やノートパソコンと直接接続することができ、情報携帯端末やパソコンが本考案の音声合成出力機能付画像合成出力装置のプログラムやデータを取り込むことができる。また、パソコンで録音・作成したデータやプログラムを本音声合成出力機能付画像合成出力装置に送って、背景パターン、音パターン、キャラクタを任意に変更することができる。

【0014】

センサ213は、赤外線センサによって、操作部位のユーザの存否を検知し、ユーザがコイン投入後、所定の位置にいるかどうか、また、存在するユーザは単数であるか複数であるかを確認する。

タッチパネル(図1のモニタ兼タッチパネル1のタッチパネル部に相当)214は、例えば、操作メニューを表示し種々の選択条件の入力を受け付けるものである。またタッチパネル214は、例えば表示した操作メニューがユーザにより入力された時の抵抗値を、CPU201がA/D変換装置を介して検知し、画面上の座標を算出することで、ユーザによりどのボタンが選択されたかを判定する。

コイン制御部215は、コインの投入の有無や不正コインの投入などをCPU201により検知する。投入されたコインの形状と重さを計測し、真正コインではないコインに対し、さらにコインの表裏の模様を判定し、投入されたコインが真正コインか、不正コインや偽造コインであるかどうかを判別する。撮影料金はコイン制御部215が受け付けるが、料金受納手段はコイン検知器に限られず、紙幣読取器、プリペイドカード読取器、クレジットカード等を用いることも可能である。

タイマ216は、外部にある管理サーバと専用線で接続できる環境にないときに装置本体内に設けられ、CPU201の命令により動作する。

【0015】

ところで、ユーザの撮影はコイン投入後、ユーザへの撮影指示をタッチパネル214が受け付けたときに、カメラ(図1のカメラ6に相当)220が行なう。画像入力装置としては、カメラ220にはCCDカメラのほかにMOSや通常のビデオカメラ等を用いてもよい。カメラ220が撮影した画像データは、カメラ220が内蔵するAE(オートアイリス)機能が露出調整し、メモリ202に記憶することもできる。

また画像入力装置として例えば、Ir-D A光通信機能を有するデジタルスティルカメラを用いてもよい。すると、あらかじめ撮影した画像データを受信し転送できるIr-D A光通信機能と複数枚の画像データが格納できるバッファメモリにより、ユーザにメニュー画像の選択条件となるアイコンを選択させて、JPEG圧縮した複数枚の静止画を容易に合成することができるようになる。

【0016】

一方、音声や背景音などの録音は、例えば、ユーザにより声を録音するコマンドや背景音を録音するコマンド等が選択され、タッチパネルがそのコマンドを受け付けることにより行われる。音声入力装置としてはマイク221のほかに、外部機器接続用LINEOUT222に接続するレコーダやスピーカ等でもよい。出力について言えば、画像・音声出力部207がデータをビデオプリンタ217に出力した場合、ビデオプリンタ217はカラー画像を葉書もしくはシールに印刷、出力する。READ/WRITE装置218に出力した場合、READ/WRITE装置218は携帯記憶媒体219に生成した画像データや音データをデジタル情報のまま記憶する。

【0017】

なお、上記の説明では、画像・音声入力部204や画像・音声処理部205のように、便宜上画像データと音データを区別無く取扱っているが、実際は画像データと音声データを区別して取扱っており、画像データに音データを合成するようなことはない。画像データに合成できるのは画像データ、音データに合成できるのは音データだけであり、音データを画像データに合成したい時には画像データに変換してから行なう。その際のデータ変換機能は、例えば画像・音声処理手段205に備えればよいし、また独立した一手段として設けてもよい。

【0018】

また、図1の駆動装置5及び7の制御部については、図2には示していないが、I/O制御部203に接続された各駆動制御部を介して図1の駆動装置5及び7の駆動制御が行われる。また、回転台4に回転開始、停止は、例えば回転台床面に設けたフットスイッチをユーザが操作することによって指示される。カメラ6の上下動についても同様に、タッチパネル1に設けた指示スイッチ、あるいは回転台床面に設けたフットスイッチをユーザが操作することによって指示される。

【0019】

図3は、本考案にかかる音声合成出力機能付画像合成出力装置が携帯記憶媒体に出力を行なう場合の具体的な構成を示す概略斜視図である。図で示すとおり、ICカード等の記憶媒体を出し入れするための出入口301が筐体300の操作面側に設けられている。これにより、図1で示したREAD/WRITE装置11が、携帯記憶媒体内の画像データや音データの入出力を行なう事ができる。

【0020】

図4は、音声合成出力機能付画像合成出力装置の出力データの構造と出力先を示す図である。出力データ400は大きくプログラム部410とデータ部420の2つに分けられる。プログラム部410は画像データ表示機能411と音声データ再生機能412を持っており、データ部420は合成後画像データ421と合成後音声データ422の2つに分けられる。出力データ400がプログラム部410を持っているのは、出力データ400を携帯情報端末やパソコン等で読み込んだ時、あらかじめパソコンに表示再生プログラムをインストールしておくことなしに、データ部420を表示再生を可能にするためであり、画像データ表示機能411は合成後画像データ421を、音声データ再生機能412は合成後音声データ422をそれぞれ表示再生する。

【0021】

出力データ400をシール430に出力する場合は、合成後画像データ421をシール印刷機によってシール430に印刷出力する。ここで音声データを利用する場合には、音声データはあらかじめ合成後画像データ421に変換されてい

る必要がある。つまり、シール印刷は画像データしか取扱えないため、音声データは画像データに変換される必要がある。その方法としては様々な形態が考えられ、例えば、公知技術である音声を音声再生用バーコードに変換する手段を、本音声合成出力機能付画像合成出力装置に備えればよい。シール印刷を行なうと画像データとともに音声再生バーコードも印刷されるので、このバーコードを専用の音声再生バーコードリーダーで読取れば音声再生される。また、ユーザが吹き込んだ音声であれば、その音声に対して音声認識処理を行なって文字データに変換し、さらにその文字データを「フキダシ」として画像データに加工する。作成された「フキダシ」は画像パターンデータとして格納し、それを撮影画像データと合成すれば、シールに印刷した際にユーザの音声も画像データとともに印刷することができる。また、あらかじめ本音声合成出力機能付画像合成出力装置に格納されている音声パターンデータを利用するならば、その音声パターンデータの文字データも格納しておき、それを「フキダシ」として利用する等が考えられる。

【0022】

また、音声音楽や背景音の場合には、その音声データがMIDIデータであるのならば、五線譜の画像データに変換して出力することが可能である。またそのような音楽に対応づけて、曲のアーティスト名や詳細情報を格納しておき、印刷の際には画像データとして出力してもよい。このように音声データを利用する場合には、音声データを文字データや画像データに変換してから、合成後画像データ421に合成する。その後、合成後画像データ421をシール印刷機430において出力、印刷する。

【0023】

出力データ400をICカードやスマートメディアなどの携帯記憶媒体440に出力する場合は、出力データ400をそのまま携帯記憶媒体440に格納すればよい。もし、本音声合成出力機能付画像合成出力装置が2種類以上の記憶媒体出力手段を有している時は記憶媒体の種類をユーザに1つ、あるいは複数指定させてから出力を行う。携帯記憶媒体440には出力データ400がデジタルデータのままで格納されているので、ユーザはその携帯記憶媒体が使用可能な表

示再生装置があれば、いつでも、何度でも、出力データ400を再生することができる。特にプログラム部410を備えたことで、表示再生プログラムが表示再生装置にインストールされていなくても、データ部420を表示再生することが可能である。さらに、表示再生装置が画像プリント装置を備えていれば、何度でも焼き増しすることができる。また、本音声合成出力機能付画像合成出力装置においては、出力したデータにさらに加工・合成を加えて出力したり、そのデータの一部を画像パターンデータや音声パターンデータとして利用して新たに出力データを作成するなど、ユーザが任意の編集を行なうことができる。

【0024】

出力データ400をネットワーク経由で配信する場合は、ユーザに配信先メールアドレスを指定させ、その指定アドレスに出力データ400をそのまま配信する。ユーザ指定のメールアドレスには出力データ400がデジタルデータのまま送られ、先に述べたように、何度でも焼き増ししたり、ユーザが任意の編集を行なうなどはもちろん可能である。また、携帯記憶媒体の場合とは異なり、ネットワーク配信は媒体の影響を受けないので、ユーザ側の端末装置に出力データの表示再生プログラムが備わっているか、プログラム部410が作動できる環境であれば、ユーザは端末装置でいつでも出力データ400を表示・再生・加工・編集等を行なうことができる。さらに、ネットワークを経由して、ユーザの端末側から本音声合成出力機能付画像合成出力装置に画像データや音声データを送信することが可能である。それにより、本音声合成出力機能付画像合成出力装置は、ユーザ側のデータをシールやICカードなどの各種媒体に出力したり、そのデータを上書き加工して新たな出力データを作成したり、新たな出力データを作成する際に用いる画像パターンデータや音声パターンデータとして用いることができる。

【0025】

次に、いくつかの実施の形態に即して説明する。

図5は、本音声合成出力機能付画像合成出力装置を「星占い」に適用した場合の処理を説明するためのフローチャートである。以下、図を参照しながら処理の一例を説明する。最初に、カメラ220でユーザを撮影し、それを撮影画像デー

タとする（ステップ501）。「背景を合成する」や「色を変える」などのタッチパネルのコマンド入力を通じて、ユーザと本音声合成出力機能付画像合成出力装置が対話式で撮影画像データを加工していくことで、撮影画像データに画像パターンデータを合成し、ユーザにとって適切な撮影画像を作成する（ステップ502）。次に、ユーザによる生年月日を入力を受け付ける（ステップ503）。

【0026】

本音声合成出力機能付画像合成出力装置は、入力された生年月日で星占いのプログラムを実行し、「今日の運勢」を表す言葉を求める（ステップ504）。なお星占いの結果がたとえ悪い運勢でも、「今日の運勢」を表す言葉は肯定的な表現とする。例えば、「明けない夜はない。もう少し頑張ろうよ。」といったやわらかな表現を用いる。次に、求めた「今日の運勢」を表す言葉をフキダシ画像データに合成し、ステップ502で作成した撮影画像とフキダシ画像データとを合成し、合成後画像データとする（ステップ505）。ここで、フキダシ画像データとは、フキダシに「今日の運勢」の言葉が書き込まれた画像パターンデータである。次に、あらかじめ本音声合成出力機能付画像合成出力装置に用意された音声データの中から、「今日の運勢」の言葉に該当する音声パターンデータを取り出し、「今日の運勢」の言葉の音声データとする（ステップ506）。そして、背景音／背景音楽などをユーザに選択させる（ステップ507）。選択された背景音／背景音楽の音パターンデータを取り出し、それを「今日の運勢」の言葉の音声パターンデータと合成し、合成後音声データとする（ステップ508）。そして、出力データをシールに出力するかをユーザに選択させる（ステップ509）。ユーザによって出力する選択がなされた場合には、合成後画像データをシール印刷機を用いてシールに印刷する（ステップ510）。次に、出力データをICカードに出力するかをユーザに選択させる（ステップ511）。ユーザによって出力する選択がなされた場合には、合成後画像データ、合成後音声データ、表示再生プログラムをICカードに出力する（ステップ512）。最後に、出力データをネットワークで配信するかをユーザに選択させる（ステップ513）。ユーザによって配信する選択がなされた場合には、配信するメールアドレスをユーザに指定してさせ（ステップ514）、合成後画像データ、合成後音声データ、表示

再生プログラムを、そのユーザ指定のメールアドレスに配信する（ステップ515）。

【0027】

図6、図7及び図8は、図5に示した本音声合成出力機能付画像合成出力装置におけるアプリケーション「星占い」の実施の形態の出力例である。

図6は、出力データをシールに印刷した一出力例である。シール600は合成後画像データをもとに作成される。合成後画像データは撮影画像601とフキダシ602によって構成されている。撮影画像601はユーザが撮影画像データをタッチパネルのコマンド入力操作を通じて、本音声合成出力機能付画像合成出力装置と対話式に加工・合成していくことで得られた撮影画像であって、図の背景にある太陽や雲もその過程でユーザによって加えられたものである。また、フキダシ602は星占いにより得られた「今日の運勢」を表す言葉を、フキダシ画像データとして合成したものである。

図7は、出力データをICカードに出力した一出力例である。ICカード700は合成後画像データ701と合成後音声データ702と表示再生プログラム703を格納している。ここで、合成後音声データ702には、「今日の運勢」の言葉に該当する音声データ以外に、ユーザ自身が選択した背景音／背景音楽なども格納している。

図8は、上記の出力例の応用例であって、ICカード801にシール802を貼り付けたものである。これによりICカード801の内容がシール802を見ることにより瞬時に判断できるという効果がある。シール802の貼り付けは機械で自動的に行ってもよいし、ICカード801とシール802とを別々に出力し、ユーザの必要に応じてユーザ自身で貼ってもらってもよい。

【0028】

図9は、本音声合成出力機能付画像合成出力装置を「恋人／友人へのメッセージ」に適用した場合の処理を説明するためのフローチャートである。以下、図を参照して処理の一例を説明する。最初にカメラ220でユーザを撮影し、それを撮影画像データとする（ステップ901）。ユーザのタッチパネルのコマンド入力を通じて、本音声合成出力機能付画像合成出力装置が対話式で撮影画像データ

を加工し、ユーザにとって適切な撮影画像を作成する（ステップ902）。次に、ユーザがマイクから吹き込んだ「恋人／友人へのメッセージ」を受付けて、録音音声データとする（ステップ903）。吹き込まれた録音音声データをスピーカで再生し、その内容の確認をユーザに問合せOKであれば次に進み、そうでなければメッセージの吹き込みをやり直す（ステップ904）。次に、本音声合成出力機能付画像合成出力装置は録音音声データであるユーザのメッセージの音声認識処理を行い、メッセージを文字データに変換した後、フキダシ画像データとして合成する（ステップ905）。そして、ステップ902で作成した撮影画像とフキダシ画像データを合成し、それを合成後画像データとする（ステップ906）。

【0029】

次に、背景音／背景音楽などをユーザに選択させる（ステップ907）。選択された背景音／背景音楽などと、ユーザが吹き込んだ録音音声データであるメッセージとを合成し、合成後音声データとする（ステップ908）。そして、出力データをシール印刷機に出力するかをユーザに選択させる（ステップ909）。

【0030】

ユーザによって出力する選択がなされた場合には、合成後画像データをシールに印刷する（ステップ910）。次に、出力データをICカードに出力するかをユーザに選択させる（ステップ911）。ユーザによって出力する選択がなされた場合には、合成後画像データ、合成後音声データ、表示再生プログラムをICカードに出力する（ステップ912）。最後に、出力データをネットワークで配信するかをユーザに選択させる（ステップ913）。ユーザによって配信する選択がなされた場合には、配信するメールアドレスをユーザに指定してさせ（ステップ914）、合成後画像データ、合成後音声データ、表示再生プログラムを、そのユーザ指定のメールアドレスに配信する（ステップ915）。

【0031】

図10は、図9に示した本音声合成出力機能付画像合成出力装置における「恋人／友人へのメッセージ」の実施の形態の一出力例である。本出力例では、シール1000がICカード1001に貼られている。撮影画像1002はユーザが

もとの撮影画像データに加工して得られた撮影画像であって、図の背景にある星もその過程でユーザによって加えられたものである。またフキダシ1003は恋人や友人のためにユーザが吹き込んだメッセージを、音声認識によって文字データに変換した後、フキダシ画像データに合成したものである。ICカード1001は合成後画像データと合成後音声データと表示再生プログラムを格納している。

ここで、合成後音声データには、ユーザが吹き込んだメッセージ以外に、ユーザ自身が選択した背景音／背景音楽なども格納している。

【0032】

図11は、本音声合成出力機能付画像合成出力装置を「ペットと写す」に適用した場合の処理を説明するためのフローチャートである。以下、図を参照して処理の一例を説明する。最初にカメラ220でユーザを撮影し、それを撮影画像データとする（ステップ1101）。ユーザのタッチパネルのコマンド入力を通じて、本音声合成出力機能付画像合成出力装置が対話式で撮影画像データを加工し、ユーザにとって適切な撮影画像を作成し、その撮影画像をユーザ画像データとする（ステップ1102）。次に、ユーザがマイクで吹き込んだメッセージを録音音声データとして記憶し（ステップ1103）、ユーザが選択した背景音／背景音楽などを音声パターンデータから取り出す（ステップ1104）。このメッセージと背景／背景音楽を合成し、ユーザ音声データとする（ステップ1105）。次に、ユーザが挿入したICカードを受け付ける（ステップ1106）。そしてICカード内の合成後画像データと合成後音声データを取り出し、それぞれペット画像データ、ペット音声データとする（ステップ1107、ステップ1108）。このICカードは、撮影以前にユーザが自分のペットを本音声合成出力機能付画像合成出力装置において撮影したものである。つまり、ICカード内の合成後画像データにはペットの画像が、合成後音声データにはペットの音声が、それぞれ格納されている。次に、ユーザ画像データとペット画像データを合成し、それを合成後画像データとする（ステップ1109）。さらに、ユーザ音声データとペット音声データを合成し、それを合成後音声データとする（ステップ1110）。そして、出力データをシール印刷機に出力するかをユーザに選択させ

る（ステップ1111）。ユーザによって出力する選択がなされた場合には、合成後画像データをシール印刷機に出力する（ステップ1112）。次に、出力データをICカードに出力するかをユーザに選択させる（ステップ1113）。ユーザによって出力する選択がなされた場合には、合成後画像データ、合成後音声データ、表示再生プログラムをICカードに出力する（ステップ1114）。最後に、出力データをネットワークで配信するかをユーザに選択させる（ステップ1115）。ユーザによって配信する選択がなされた場合には、配信するメールアドレスをユーザに指定してさせ（ステップ1116）、合成後画像データ、合成後音声データ、表示再生プログラムを、そのユーザ指定のメールアドレスに配信する（ステップ1117）。

【0033】

図12は、図11に示したアプリケーション「ペットと写す」の実施の形態において、ユーザが本音声合成出力機能付画像合成出力装置に挿入したICカードである。

ICカード1200には、合成後画像データ1201、合成後音声データ1202、表示再生プログラム1203が格納されている。合成後画像データ1201は、ペットの猫の撮影画像である。合成後音声データ1202は、その猫の「ミャオミャオ」という鳴き声である。

【0034】

図13は、図11に示した本音声合成出力機能付画像合成出力装置のアプリケーション「ペットと写る」の実施の形態における、ICカードへの出力例である。使用するペットのデータは図12に示したICカードから取り込んでいる。ICカード1300には、合成後画像データ1301、合成後音声データ1302、表示再生プログラム1303が格納されている。合成後画像データ1301は、ユーザの撮影画像に、ICカードの猫の撮影画像を合成したものである。合成後音声データ1302には、ユーザの「今度遊びに来て下さいね」というメッセージとともに、ICカードから取り込んだ猫の「ミャオミャオ」という鳴き声も格納されている。

【0035】

図11、図12及び図13で示した実施の形態では、ICカードから合成用のデータを取り込んでいるが、他の携帯記憶媒体でもよいことはいうまでもない。例えば、デジタルカメラ内の携帯記憶媒体から撮影したデータを読み込んで、合成するための画像データとすることができる。逆に、携帯記憶媒体内の画像データ及び音声データを読み込んで、その上に上書き加工、合成をすることも可能である。

【0036】

図14は、本音声合成出力機能付画像合成出力装置を「コマ割りドラマ」に適用した場合の処理を説明するためのフローチャートである。

一般にカラオケというものは音楽で歌手の歌う部分をカラにし、ユーザが自分の歌声でその部分を補うことによって素人が歌手になった気分を幾分味わえるようにしたものである。それに対してここで述べるコマ割りドラマは、静止画ではあるが、ドラマでヒロイン（またはヒーロー）の演じる部分をカラにし、ユーザの肖像及び声でその部分を補うことによってユーザがドラマのヒロイン（またはヒーロー）になった気分を幾分味わえるようにしたものであるとすることができる。その際ユーザの相手役としては人気俳優、アニメのキャラクタ等が考えられる。

【0037】

以下、図15のコマ画像遷移図を参照しながら、図14のフローチャートの処理の一例を説明する。

最初に、あらかじめ用意された複数のコマ割りドラマの中から好みのコマ割りドラマと、そのドラマに出てくるキャラクタのうち演じたいキャラクタを、ユーザにタッチパネルで選択させる（ステップ1401）。コマ割りドラマのコマ数はあらかじめ決められていてもよいし、撮影が始まる前にユーザに指定してもらってもよい。コマ数が増加する場合は、コマ数に応じて課金を要求する。そしてそのユーザが選択したドラマの最初のコマを読み出す（ステップ1402）。次に、コマを表示する前にコマの中からユーザが選択したキャラクタを消去し、そのキャラクタが表示されていた場所は他と区別して画面に表示する（ステップ1403）。この区別して表示した場所にユーザの撮影画像をはめ込む。図15の

(A) がもとのコマ画像、(B) がステップ1403の処理済みコマ画像である。 (A) の1501はユーザが選択したユーザ選択キャラクタであり、このキャラクタ画像の代わりにユーザ画像をはめ込む。ステップ1403の処理でユーザ選択キャラクタ1501は消去され、その位置には他と区別して表示するために、例えば図のように網掛け1502が行われている。この網掛け部分1502にユーザがはめ込まれる。ユーザが表示された処理済みコマ画像(B)を見ながら好きなポーズをとり、カメラで撮影を行ない撮影画像データとする(ステップ1404)。この時ユーザは回転台を回転させたりカメラの上下を操作したりして、好きなアングルで撮影することができるようにも良い。そして、撮影画像データを図15の処理済みコマ画像(B)と合成する(ステップ1405)。その結果が図15の(C)で示された合成後コマ画像である。網掛け部分1502には、ユーザ選択キャラクタの代わりにユーザの画像1503がはめ込まれ表示されている。次は、ユーザにキャラクタの会話を指示する。まず、現在撮影中のコマのユーザ担当キャラクタのセリフを表示する(ステップ1406)。図15の例では、「明日は晴れるかしら」というセリフが画面に表示されることになる。そして、ユーザはその表示に従ってセリフを吹き込み、その音声データを格納する(ステップ1407)。そして、コマ内の会話に必要な各キャラクタの音声データを読み出し、その音声データを格納する(ステップ1408)。ちなみに各キャラクタの会話やセリフの音声データは、あらかじめ記録されている。たとえば、図15の相手役キャラクターの「多分晴れると思うよ」というセリフは、あらかじめ音声パターンデータとして記録されている。さらに、各キャラクタのセリフの再生順序は、再生した時に自然な会話に聞こえるような配列になっている。つまり、ユーザの吹き込んだセリフがユーザ選択キャラクタの音声パターンデータの代わりに格納されるので、音声を再生した時にはユーザがあたかも他のキャラクタと会話しているように聞こえる。これらの会話の音声データとBGMや効果音を合成し、合成後コマ音声とする(ステップ1409)。最後に、最後に合成後コマ画像1503を画面に表示し、ユーザの吹き込んだセリフを含めたコマ内の全ての会話を音声再生し、ユーザにこれで良いか確認する(ステップ1410)。気に入らなければステップ1404に戻り、撮影と録音をやり直す。

以上で1コマ分の作成が終わったので、次に続きのコマがあるか判定する（ステップ1411）。もし続きのコマがあれば、次のコマを読み込んでステップ1403に戻る（ステップ1412）。続きのコマが無く全てのコマの作成が終わったら、出力処理に入る。全コマ分の合成後コマ画像及び合成後コマ音声、それぞれ合成後画像データ及び合成後音声データである。まず、出力データをシールに印刷するかをユーザに選択させる（ステップ1413）。ユーザによって出力する選択がなされた場合には合成後画像データを印刷する（ステップ1414）。次に、出力データをICカードに出力するかをユーザに選択させる（ステップ1415）。ユーザによって出力する選択がなされた場合には、合成後画像データ、合成後音声データ、そして表示再生プログラムをICカードに出力する（ステップ1416）。最後に、出力データをネットワークで配信するかをユーザに選択させる（ステップ1417）。ユーザによって配信する選択がなされた場合には、配信するメールアドレスをユーザに指定させ（ステップ1418）、合成後画像データ、合成後音声データ、表示再生プログラムを、そのユーザ指定のメールアドレスに配信する（ステップ1419）。

【0038】

図16は、本音声合成出力機能付画像合成出力装置のアプリケーション「コマ割りドラマ」の実施の形態において、ICカードへの出力の一例である。ICカード1600には合成後画像データ1601と合成後音声データ1602と表示再生プログラム1603が格納されている。合成後画像データ1601は複数の合成後コマ画像からなり、この場合4コマから構成されている。そして図に示されている女性、ユーザ選択キャラクタに代わってはめ込まれているユーザである。合成後音声データ1602には、コマ割りドラマの会話が音声として格納されている。下線が引かれている部分はユーザが担当した部分、つまりユーザがマイクから吹き込んだセリフである。そして、1603が表示再生プログラムであるが、本実施の形態における表示再生プログラムとは、単に画像を表示し音声を再生するに留まらず、コマ順に合成後コマ画像を表示し、表示した各コマに対応して合成後コマ音声を再生することができるものである。

【0039】

以上の本考案の実施の形態によって、本考案が画像データだけでなく音声データも再生できることは明らかであるが、本考案による音声の再生方法は本考案の実施の形態に留まらない。以下、その他考えられる再生方法の例を図を参照して説明する。ところで、以下の出力例では音声データを音声のまま再生するので、印刷されている合成後画像データには「フキダシ」は用いていない。

【0040】

図17は、シール印刷のみで音声再生を行なった一例である。シール1700には、印刷画像1701のほかに、音声再生用バーコード1702が印刷されている。これは公知技術であるところの、音声を音声再生用バーコードに変換、印刷する技術を用いている。音声を再生する時は、図示しない専用のバーコードリーダーでシール1700に印刷された音声再生用バーコード1702を読取ればよい。

【0041】

図18は、携帯記憶媒体のみで音声再生する再生方法の一例である。ICカード1800には印刷したシール1801を貼ってあるので、画像データはシール1801を見ることによって得られる。音声データについては、ICカード1800内の音声再生素子1802を動作させることにより再生が可能である。つまりICカード1800だけで画像と音声を楽しめることになる。

【0042】

図19は、携帯端末で音声データを再生をする再生方法の一例である。キーホルダー型ICカードREADER1900は小型の携帯端末であり、画像表示手段と音声再生手段を備える。これにより、表示再生プログラムを起動し、表示部1901には合成後画像データが表示され、音声出力部1902からは合成後音声データが出力される。

【0043】

図20は、パソコン2000で合成後画像データと合成後音声データを表示再生する再生パターンの一例である。ICカードREADER2001から読み込んだ、またはネットワーク2002経由でメールとして届いた出力データ（合成後画像データ＋合成後音声データ＋表示再生プログラム）を受信し、パソコン2

000は表示再生プログラムを読取って起動し、モニタ2003及び内蔵スピーカ2004で表示、再生する。

【0044】

他の再生パターンとして、図示しないが本音声合成出力機能付画像合成出力装置そのものを利用した再生パターンがある。そのためには、本音声合成出力機能付画像合成出力装置そのものにICカードから出力データを読み込み表示、再生する機能を持たせておけばよい。

【0045】

【考案の効果】

以上説明したように本考案は、従来の画像プリント装置に画像データだけでなく、録音されたユーザの声、自然音等の背景音、背景音楽等の音データを扱える機能を加えたことにより、画像に音声を加えた、より現実に近いアミューズメント環境をユーザに提供することができる。

また、画像データと音データに表示再生プログラムを加えて、携帯記憶媒体への書き込み、ネットワーク経由での指定メールアドレスへの配信を行うことにより、ユーザは後で画像データと音データを表示、再生することができる。